UNIVERSITY OF CALIFORNIA
Santa Barbara

In Vitro Directed Evolution of Fluorescence-Enhancing and Structure-Switching RNA Aptamers

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Materials

by

Michael Roy Gotrik

Committee in Charge:

Professor H. Tom Soh, Chair

Professor Craig J. Hawker

Professor Irene A. Chen

Professor Omar A. Saleh

June 2017

The Dissertation of
Michael Roy Gotrik is approved:

Professor Craig J. Hawker

Professor Irene A. Chen

Professor Omar A. Saleh

Professor H. Tom Soh, Committee Chairperson

June 2017

In Vitro Directed Evolution of Fluorescence-Enhancing and Structure-Switching RNA Aptamers

Michael Roy Gotrik

Acknowledgements

Foremost, I would like to thank my mentor and advisor: Prof. Tom Soh. I met Tom quite by chance during an informal and unplanned session during visiting weekend at UCSB. I was immediately drawn by his engaging personality, fervor for impactful research, and the seemingly limitless potential of his research focuses. Over the last five years, I have undergone a dramatic transformation both as a scientist and as an individual, and none of it would have been possible without the advice and opportunities that Tom has granted me. Since joining the Soh Lab late in 2012, I have not once looked back.

I also owe a debt of gratitude to my other committee members: Prof. Craig Hawker, Prof. Omar Saleh, and Prof. Irene Chen. Their realism and insight into the many ideas we have discussed over the years has helped lay the groundwork of my scientific career and their support and high expectations have been essential to my success.

I would not have made it this far without a number of important mentors in my life. I would like to thank Dr. Andrzej Rajca, Professor of Chemistry at the University of Nebraska - Lincoln (UNL), and Dr. Hendrik Viljoen, Professor of Chemical Engineering at UNL. My first experiences with research were in their labs; and their patience and gentle guidance is what, in the end, has allowed me to burgeon into a full-fledged scientist. In addition, Dr. Luke Connal, now a Senior Lecturer at the University of Melbourne in Australia, was my mentor during an NSF REU internship at UCSB in 2011. He challenged me to approach seemingly intractable problems from multiple directions, and to transform bitter failures into resounding success.

Neither this dissertation nor my doctoral degree could have been accomplished without the help of Dr. Gurpreet Sekhon. Not only has he been a great friend, but he taught me everything I know about nucleic acids and was the inspiration behind many of the projects discussed in this dissertation. He rescued me from a vicious cycle of dead-end ideas, helped illuminate the path towards many experimental successes, and consistently reminded me to take every result with a hefty dose of skepticism. Without this, I would not have learned to approach my work with the thoroughness and care required for impactful research.

Dr. Andrew Csordas has been an invaluable reservoir of knowledge, support, and guidance during my graduate career. He has always been willing to provide a new insight or offer the encouragement needed to overcome any problem. Dr. JP Wang gratefully fixed all of my poor lab techniques when I first began at UCSB, and was a shining example of how a clever idea can bring about meaningful progress. Dr. Jinwen Yu assisted me in a number of projects and became a close friend during my time here.

I would like to thank my many collaborators and work associates for their insights and assistance. Dr. Saumya Saurabh has not only gone above and beyond in completing various experiments, but has provided me much invaluable advice and guidance as I look towards my future. Wes McKeithan assisted me in a number of techniques where I would have otherwise been completely lost. I thank the many members of the Soh Lab, past and present, who have helped guide me down this path and have shared many of these experiences with me. Several of you have become close friends, and I am happy to be part of such a great, diverse, and talented group of people.

My family has been a constant source of inspiration for me. My parents, Roy and Jytte Gotrik, provided me the world and everything I needed to get to where I am. My oldest brother Kevin has been an exemplary role model throughout my life and I would not be where I am without his help. His wife, Susan, and their beautiful daughters (Felicity, Eleanor, Linnea, and Ingrid) provide a constant source of unadulterated joy to all of us. My brother David consistently reminds me to follow my passions and is always there to offer a helping hand. Last but not least, my sister Jennifer has become one of my closest friends and never fails to make me jealous of her travels or boundless talent.

I also extend my most heartfelt thanks to all of my friends for their companionship, support, and role in allowing me to grow into the person I am today. The last 5 years have truly been the best of my life. In addition to those already mentioned: Michael Trogdon, Surabhi Ram, Casey Heier, Chris Lander, Stefan Heinz, Kate Hake, Humberto Foronda, Andrew Hamilton, Senay Hawelti, Anna Melker, Luis Rodriguez, Donald Molosi, Effie Greene, Tori Fry, Drew Peterson, Omri Almog, Alex Yoshikawa, Sam Vakhidov, Trent Linbo, Jimmy Duin, and Dylan Mangel. Of course, there are countless others who have offered me kindness, support, friendship, guidance, and hospitality. For that and everything else, thank you!

Curriculum Vitæ

Michael Roy Gotrik

mgotrik@gmail.com

Education

June 2017 **Doctor of Philosophy in Materials**
Bioengineering Emphasis
University of California, Santa Barbara (UCSB)

June 2012 **Bachelor of Science in Chemical Engineering**
University of Nebraska, Lincoln (UNL)

Journal Publications

3. **Gotrik, M. R.**, Sekhon, G., Saurabh, S., McKeithan, W, Nakamoto, M, Eisenstein, M., Moerner, W.E., Soh, H.T., Direct Selection of Fluorescence-Enhancing RNA Aptamers, in preparation.
2. Nothling, M. D., Ganesan, A., Condic-Jurkic, K., Pressly, E., Davalos, A., **Gotrik, M. R.**, et al., Coote, M. L. (2017). Simple Design of an Enzyme-Inspired Supported Catalyst Based on a Catalytic Triad. *Chem*, 2(5), 732-745.
1. **Gotrik, M. R.**, Feagin, T. A., Csordas, A. T., Nakamoto, M. A., Soh, H. T. (2016). Advancements in Aptamer Discovery Technologies. *Accounts of Chemical Research*, 49(9), 1903-1910.

Conference Proceedings

1. **Gotrik, M. R.**, Sekhon, G., Soh, H.T. "Direct Selection of RNA Aptamers for Fluorescence Enhancement".7th International Conference on Bioengineering and Nanotechnology. Chicago, Illinois 2017.

Poster Presentations

4. Canary Foundation Early Detection Symposium. 2017.
3. Canary Summit for Early Cancer Detection. 2016.
2. Stanford SystemX Spring 2016 Conference.
1. Center for NanoSciences (CeNS) Workshop. Venice, Italy 2015.

Activities and Honors

2016	Canary Challenge Volunteer
2015	Junior Nanotech Network (JNN) Program Participant. München, Germany
2014	Family Ultimate Science Exploration (FUSE) Volunteer
2014 - 2015	SciTrek Volunteer
2012	Honor's Program Graduate, With Distinction. University of Nebraska - Lincoln
2012	Departmental Outstanding Senior Award. Chemical Engineering Department. University of Nebraska - Lincoln
2011	NSF Research Experience for Undergraduate Participant at University of California, Santa Barbara
2011	Deutsche Akademische Austauschdienst (DAAD) Awardee (declined)
2011	Milton E. Mohr Scholarship. College of Engineering. University of Nebraska - Lincoln
2011	Global Gateway Travel Scholarship Awardee. University of Nebraska - Lincoln
2008 - 2012	Regents Scholarship Awardee. University of Nebraska - Lincoln

Abstract

In Vitro Directed Evolution of Fluorescence-Enhancing and Structure-Switching RNA Aptamers

Michael Roy Gotrik

Genetically-encoded fluorescent proteins have become an essential tool, and researchers today have access to a diverse array of protein-labeling strategies for studying fundamental *in vivo* processes. However, there exist relatively few methods for *in vivo* detection of an equally important biomolecule RNA. One useful method to track RNA entails fusing the cellular RNA of interest to an RNA aptamer that can bind to and turn on the fluorescence of a small-molecule dye. Unfortunately, it is difficult to discover such fluorescence-enhancing RNA aptamers. This is because the conventional method of aptamer discovery (SELEX) can enrich for RNA that binds to a dye, but it does not preferentially identify sequences that generate a fluorescent signal upon binding. Thus, there is a critical need for a method capable of directly screening RNA aptamers for fluorescence enhancement.

To address this need, we have developed a strategy for rapidly and intently screening large numbers of RNA aptamers based on their capacity to generate a fluorescent signal upon binding a small-molecule dye. Our approach generates libraries of Gene-linked RNA Aptamer Particles (GRAPs) that display functional

RNA aptamers and are isolatable using fluorescence-activated cell sorting (FACS). As proof of concept, we performed selections isolating fluorescence-enhancing aptamers against malachite green (MG). We show that by directly selecting for function rather than binding, we can isolate RNA aptamers that are brighter and higher affinity than the best known MG aptamer. GRAP display also enables us to measure the fluorescence signal of every aptamer across multiple emission windows, permitting us to reconstruct the emission profile of every displayed aptamer. This in turn, allows us to intentionally isolate aptamers that fluoresce at a variety of wavelengths upon binding their target dye. Lastly, this technique can be used to discover functional RNA that undergoes a structural re-organization upon ligand binding. This has promising implications in the fields of biosensing, gene therapy, and cellular computing. This flexibility should greatly expand the toolbox of reagents available for studying RNA in vivo.

Contents

List of Figures

Chapter 1

Introduction

1.1 Fluorescence Biology

Since the discovery of the green fluorescent protein (GFP) in 1962[1] and its first recombinant expression 32 years later[2,3], GFP and its many derivatives have helped revolutionize nearly all fields relating to biology. Through application with a number of technologeis, fluorescent proteins (FPs) allow researchers to visualize structures and processes in living systems that would otherwise be unobservable. For example, FPs have played key roles in determining how proteins localize in the cell[4,5], developing and studying complex intracellular networks[6,7], and

performing cellular microscopy below the diffraction limit[8]; among numerous other functions[9–11].

FPs are widely used in part because they can be expressed *in vivo* and function without requiring external catalysts or cofactors[9]. This means the proteins can be synthesized and assembled in a cell and fluorescent measurements can be made while requiring little perturbation of the expressing system. Though FPs have allowed fantastic advancements in understanding protein biology, there exist few methods of studying the well-known precursors of proteins - nucleic acids (DNA or RNA). This is because, except for the so-called "Y base" in the tRNA of yeast[12], there is no naturally fluorescent nucleic acids. In other words, no natural nucleic acid sequence, without the addition of external organic dyes, will itself assemble into a self-fluorescing system[13]. Thus, while our understanding of protein biology is constantly improving, we remain mostly "in the dark" when it comes to understanding the complex roles that DNA and RNA can play in the cell. To understand how best to address this issue, it is helpful to gain a basic understanding of the physical processes at play that causes a molecule to fluoresce.

1.2 Properties of Fluorescent Molecules

Though the earliest written account of fluorescence is as early as 1565[14], William Herschel[15] described the fluorescence phenomenon well in 1845 when he wrote:

> Equal weights of the sulphate (of quinine) and of crystallized tartaric acid, rubbed up together with addition of a very little water, dissolve entirely and immediately. It is this solution, largely diluted, which exhibits the optical phenomenon in question. Though perfectly transparent and colorless when held between the eye and the light, or a white object, it yet exhibits in certain aspects, and under certain incidences of the light, an extremely vivid and beautiful celestial blue colour, which, from the circumstances of its occurrence, would seem to originate in those strata which the light first penetrates in entering the liquid.

The event that Herschel described was that of quinine becoming excited by sunlight and yielding a "celestial blue" fluorescence emission. Though inexplicable at the time, the mechanism of fluorescence has since become a well-established theory.

Simply put, fluorescence is the result of an electron in a molecule entering an excited singlet state due to absorption of light, and then returning to its ground (unexcited) state with the emission of a photon. This process is rapid (a typical lifetime being around 10 nanoseconds)[16], yet can vary wildly depending on the chemical properties of the excited compound. While many materials are known to exhibit fluorescence (e.g. quantum dots, divalent manganese or europium salts, FPs) the most common class of fluorescent molecules are organic. This is because organic, aromatic compounds that contain large delocalized π systems can strongly

absorb incoming light. Thus, many electrons enter the excited singlet state which is a necessary precursor for fluorescence. These aromatic compounds, often called chromophores due to their propensity to absorb in the visible range, are then able to release the absorbed energy through radiative (i.e. fluorescence) or non-radiative (i.e. thermal) processes. To transform a light-absorbing chromophore into a light-emitting fluorophore simply involves the reduction of non-radiative decay channels in favor of radiative channels. This can be achieved through covalent or steric restriction of rotational and vibrational modes.

Chromophores that fluoresce upon steric restriction of non-radiative decay channels form an interesting class of molecules termed "molecular rotors[17]." Usually, these chromophores are non-fluorescent when free in solution but are "turned on" when their intramolecular rotations become constrained by an outside actor. This process, aptly termed "restriction of intramolecular rotation[18]" (RIR), is widely exploited in the study of materials that undergo aggregation-induced emission (AIE)[19]. In this process, a chromophore will self-aggregate upon precipitation from a solution at which point fluorescence becomes its primary decay mechanism.

RIR can also be induced through specific interaction with a given biomolecule. These types of interactions are possible with so-called "affinity reagents" which bind tightly and specifically to their target molecule. For example, a collection of "blue-fluorescent antibodies" were developed[20] that activate derivatives of stilbene

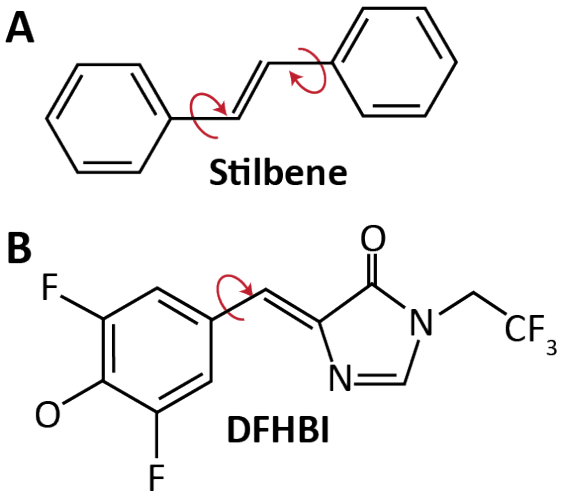

Figure 1.1: Chemical structures for stilbene (**A**) and 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI, **B**). The rotation modes that must be restrained in each molecule in order to activate fluorescence are indicated by red arrows.

(Figure 1.1 A). Likewise, the well-known Spinach RNA aptamer interacts with and activates 3,5-difluoro- 4-hydroxybenzylidene imidazolinone (DFHBI, Figure 1.1 B). This demonstrates how non-fluorescent biomolecules can be made to fluoresce by interacting with and "turning on" an external chromophore.

1.3 Aptamers as Affinity Reagents

Beyond their role in enhancing the fluorescence of chromophores, affinity reagents that bind tightly and specifically to a target molecule are an indispensable tool in biological and medical research, diagnostic tests, pharmaceuticals, and targeted therapies. Indeed, virtually every field of research in biomedicine relies to some

degree on being able to recognize and act upon a molecule of interest. Antibodies are most commonly used in this role, owing to their immense chemical diversity and functionality, excellent recognition abilities, and the enormous infrastructure in place for discovering, characterizing, and disseminating these reagents. Indeed, as of 2011, there was estimated to be over 500,000 unique antibodies towards over 15,000 targets available to researchers[21,22] worldwide.

Because antibodies are isolated through exploitation of a live animal's immune system, their discovery is inherently costly and it can be difficult to identify antibodies against toxic, unstable, self-similar, or non-immunogenic targets[23]. Oftentimes, the resulting antibodies can themselves exhibit immunogenicity and thus be unsuitable for *invivo* and medical treatment applications[24,25] In addition to these problems, numerous large-scale antibody validation studies have shown that there exists widespread issues with the reported specificity and reproducibility of commercially-available antibodies[26–28]. For example, an antibody validation project reported in 2007 that of nearly 18,000 antibodies characterized, only about 35% were shown to work for their intended purposes[21]. For these reasons and others, there is growing interest in affinity reagents that can perform the same functions as antibodies while being inexpensive, reproducible, and easily disseminated[29–34].

Aptamers - whose names is derived from the Latin *aptus* meaning 'to fit' - are short oligonucleotides that bind to or "fit" their target molecule. Aptamers can perform many of the same functions as antibodies with a number of important differences. Whereas antibodies gain their extensive chemical functionality through utilization of the 20 natural amino acids; nucleic acid aptamers gain their properties from their sugar backbone (ribose in RNA, deoxyribose in DNA) and combinations of the 5 nucleobases - adenine (A), guanine (G), cytosine (C), thymine (T, DNA-specific), and uracil (U, RNA-specific). By adopting complex 3-dimensional structures and displaying their nucleobases in precise spatial orientations, aptamers bind their targets through non-covalent interactions such as hydrogen bonding, charge-charge interactions, and Van der Waals forces[35]. While aptamers are generally considered to bind less-tightly (having low affinity), less-specifically, and to fewer potential targets than antibodies; many aptamers have been discovered with picomolar affinity[36,37]; unparalleled specificity[38,39]; and against targets such as proteins[40], small molecules[41], whole cells[42,43], and more[44,45]. One common example of these capabilities is an aptamer against the small molecule theophylline which has a >10,000-fold reduced affinity to caffeine on the basis of a single methyl group difference[38].

Discovered concurrently in the labs of Larry Gold[46] and Jack Szostak[47] in 1990, aptamers have since been shown to hold incredible promise to supplement or

otherwise replace antibodies in many applications[29,48]. This is because antibodies are proteins discovered through the exploitation of a live animal's immune system and thus naturally possess a number of disadvantages when compared to aptamers. For example, while antibodies must be produced by living systems and distributed as a physical specimen, nucleic acid aptamers can be reproducibly synthesized using well-established chemistries and can be shared digitally as sequence information. This greatly reduces the per-mole cost of aptamers compared to antibodies and lends to their unparalleled reproducibility[25,49].

The purely synthetic nature of aptamers makes them amenable to chemical and structural modifications in order to improve their affinity, specificity, and chemical or thermal stability[29,50]. They can be easily modified to include fluorophores for use as biosensors[51–53] and unnatural backbones, sugars, and nucleobases to improve function and nuclease resistance *in vivo*[54–56]. In addition, aptamers are generally thermo-stable and, if they become improperly folded at any time during their application, they can often be correctly refolded by simple temperature cycling - a capability rarely allowed by proteins. This means aptamers are more tolerant to prolonged storage and environmental variability to which antibodies are highly susceptible[27,57]. Lastly, aptamers are mostly discovered through a process of *in vitro* directed evolution meaning they can address a host of targets

that are incompatible with *in vivo* antibody discovery - including highly toxic, non-immunogenic, and biologically-labile targets[29].

1.4 Systematic Evolution of Ligands by Exponential Enrichment (SELEX)

To understand the advantages that aptamers hold and the challenges they have yet to overcome, it is necessary to be cognizant of the method by which they are discovered. The process, termed Systematic Evolution of Ligands by EXponential enrichment[47] (SELEX) begins with a large diverse "library" of random sequences. Though it is possible to derive libraries from genomic DNA[58] or transcriptomic sources[59] and using modified nucleotides[60–63], the most common method is to produce the library using phosphoramidite chemistry. During the synthesis process, the library is designed to incorporate 20≈80 "random" bases; or bases that have an equal probability to be any of the four nucleobases.(Figure 1.2 A). This random region is surrounded on either side by constant primer binding sites which are necessary to amplify and manipulate the library using polymerase chain reaction (PCR), transcription, and other enzymatic methods of synthesis, replication, or digestion. Because each random base is one of four possibilities (A,T/U,C, or G),

the total library diversity scales as 4^n where n is the total number of random bases. Thus, the theoretical diversity of a typical library is generally $10^{12} \approx 10^{48}$.

Because of physical restraints on the amount of DNA that can be feasibly used, each round of SELEX begins with a pool containing around 10^{14} sequences (<1 nanomole). This pool undergoes a cyclical process (Figure 1.2 B) wherein it is first exposed to the molecule of interest, target-bound sequences are separated from unbound sequences (which are discarded), and bound sequences are recovered for use in further rounds of SELEX. If DNA is used in this process, the recovered sequences are PCR amplified and the antisense strand is enzymatically digested to regenerate the pool. If performed using RNA, collected sequences are first reverse transcribed into DNA, PCR amplified, and then transcribed into RNA to start the next round. Though dozens of variants of SELEX have been described[64], they often rely on these basic principles. Typically, over the course of 6-20 rounds, the starting pool goes from containing $\approx 10^{14}$ unique sequences, to being primarily made up of a small number of unique sequences that recognize and bind to the target molecule. The rapidity and magnitude of pool *convergence* (reduction of pool diversity) oftentimes determines whether a selection is deemed a success or failure. If no convergence occurs, suitable candidates cannot be identified.

It is important during SELEX to monitor when and how quickly convergence occurs, as it is desirable to use the fewest number of rounds in order to minimize

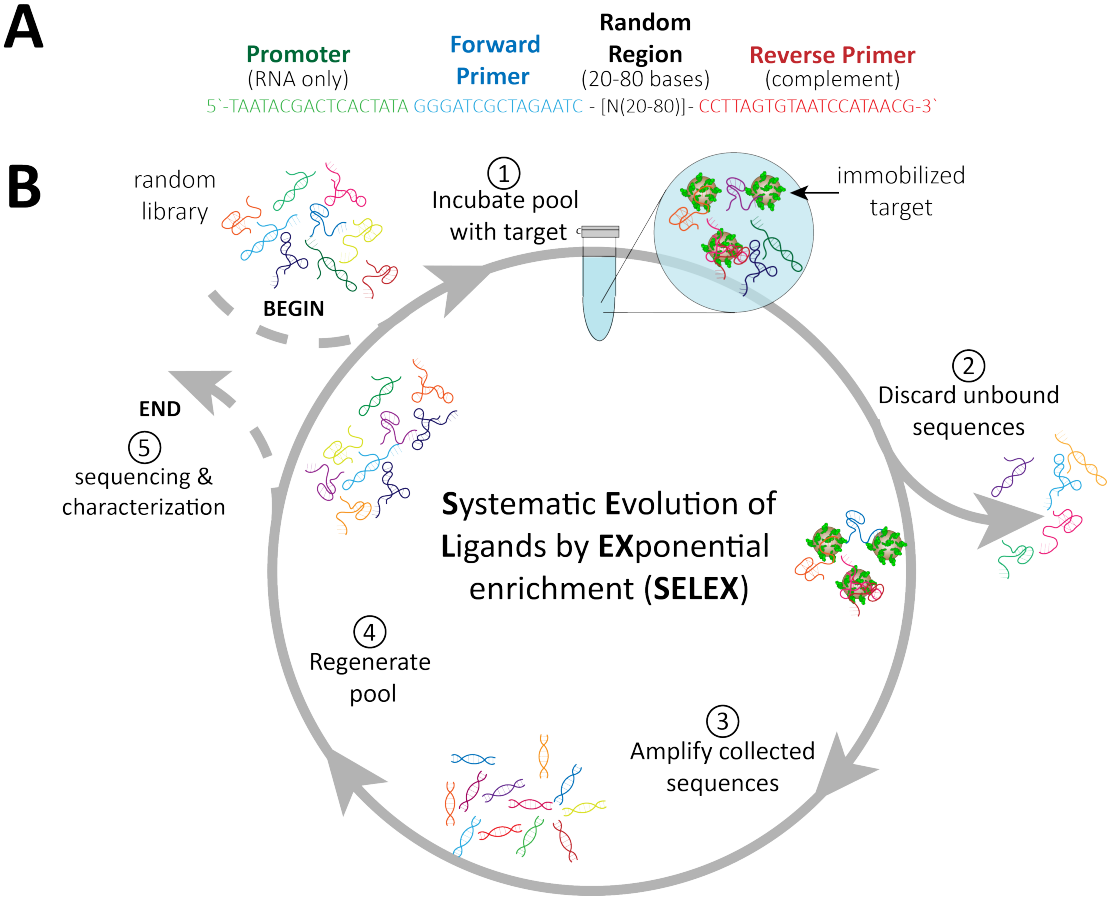

Figure 1.2: Systematic Evolution of Ligands by EXponential Enrichment (SELEX) Process. **A)** Typical library design for use in directed evolution. RNA polymerase promoter region (green) is only necessary when performing SELEX using an RNA library. **B)** Selection scheme for SELEX. **Step 1**: A random pool of single-stranded nucleic acids is incubated with an immobilized target. **Step 2**: Unbound sequences are removed and discarded. **Step 3**: Bound sequences are reverse-transcribed (RNA only) and amplified using PCR. **Step 4**: Double-stranded DNA is regenerated into a single-stranded aptamer pool through exonuclease digestion (DNA) or reverse-transcription (RNA). Steps 1-4 are typically repeated 6-20+ times until pool converges to a handful of functional sequences. **Step 5**: Pool undergoes sequencing and candidate aptamers are characterized.

time, resources, and biases (see Section 1.6) inherent to the process. Convergence of a pool is sometimes monitored by measuring the amount of recovered nucleic acids (e.g. using qPCR[65] or a radio-labeled library[66]) or the bulk binding affinity of the library (e.g. using fluorescence-activated cell sorting (FACS)[67], filter binding[68], or capillary electrophoresis[69]). When the pool is measured or shows indications of convergence, it undergoes sequencing in order to identify potential aptamer candidates. Upon sequence analysis, suitable candidates are synthesized and characterized in order to identify the best-performing aptamers. Though some ambiguity exists in the exact meaning of the term "aptamer", in this dissertation it refers to nucleic acid sequences that exhibit function (e.g. binding, catalysis, switching, etc.) and are designed or discovered using directed evolution (e.g. SELEX).

1.5 Functional Nucleic Acids and Their Application

While DNA and RNA are best known as messengers of genetic information, researchers are constantly identifying new and interesting functions of these nucleic acids in nature. While it was once believed that 99% of the genome (that is, the non-protein-coding regions) was simply "junk" DNA, this idea has largely fallen wayside

as researchers continue to find new and interesting biological and biochemical roles of these non-coding nucleic acids *in vivo*. For example, the Encyclopedia of DNA Elements (ENCODE) Project[70], announced in 2012, estimates that >75% of the human genome is biochemically active and transcribed into non-coding RNA serving a variety of purposes[71].

Functional RNA is, in fact, ubiquitous in cellular biology. In the ribosome, for instance, protein plays a primarily structural role[72]; while ribosomal RNA (rRNA) is responsible for reading messenger RNA (mRNA), aligning transfer RNA (tRNA) to the correct mRNA codon, and catalyzing the transfer of amino acids onto the peptide chain of a protein undergoing translation[73]. Encouraged by the catalytic potential of nucleic acids, large numbers of ribozymes (RNA-based) and deoxyribozymes (DNA-based, also called DNAzymes) have been evolved and characterized over the last 20 years spanning a multitude of catalytic reactions both chemical and biochemical[74–76]. Together, these discovered catalytic nucleic acids are often called aptazymes - a portmanteau of aptamer and enzyme. Aside from catalytic roles, non-coding RNAs are also found to play roles in subcellular structural organization[77,78]; promote gene silencing[79,80], gene expression[81,82], and stem cell differentiation[83]; determine macromolecular localization[84]; and many other functions[81,85,86].

1.5.1 Nucleic Acid Switches

While aptamers do not currently possess the depth nor range of applications performed by the various functional nucleic acids *in vivo*, aptamers have been discovered that can perform many interesting functions that other affinity reagents are yet incapable of addressing. For example, many aptamers can fold reversibly and specifically in the presence of their target and find use as molecular sensors[87]. This folding is often a result of intermolecular interactions whereby a separate, complimentary nucleic acid interacts with the aptamer until introduction of an external ligand that results in an energetically more-favorable folding state. A measurable change in structure is then coupled to arbitrary downstream processes and can yield a variety of applications such as real-time small molecule[88,89], metal ion[90], and protein detection[91,92]. For example, a structure-switching DNA aptamer has been utilized to measure[93] and then control[94] the levels of an important cancer drug - doxorubicin - in live animals. This was done by coupling a structure-switching aptamer to an electrode inside a microfluidic chip and monitoring folding through an electronic signaling cascade which is directly correlated to a real-time target concentration.

Riboswitches are a specific class of structure-switching RNA that are commonly found in bacteria and viruses as one manner of regulating gene expression[95,96](see

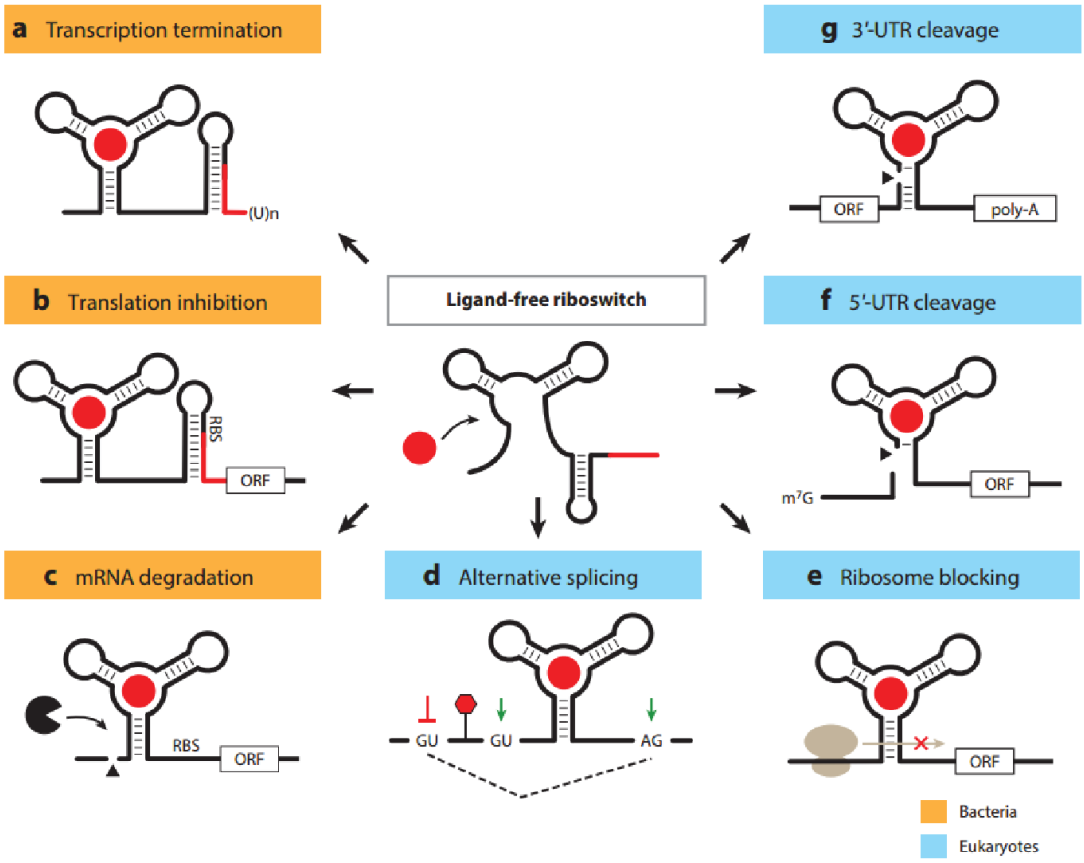

Figure 1.3: Mechanisms of gene regulation for natural and synthetic riboswitches. In orange are mechanisms used by bacteria, while blue shows eukaryotic (primarily yeast) mechanisms. **a)** In the ligand-bound form, a terminator hairpin causes transcription termination. **b)** A structural change prevents occupation of the ribosome-binding site (RBS) causing translation inhibition upstream of the open reading frame (ORF). **c)** Destabilization of folded RNA leads to rapid ribozyme degradation. **d)** Splice sites are reorganized through a structural change, leading to an alternate form of the mature mRNA. **e)** Translation start site is altered so as to be unrecognizable by the ribosome. **f)** 5′-7- methylguanosine cap is cleaved by ribozyme to prevent ribosome progression. **g)** Removal of poly-A tail in eukaryotes by ribozyme leads to rapid mRNA degradation. Reproduced with permissions [95].

Figure 1.3). They work by affecting gene expression through conformational change but only when in the presence of an external stimuli - typically a small molecule[96–99]. Riboswitches are a particularly promising topic in the field of RNA aptamers, as successful design and implementation can allow functional expression of 'RNA machines' or genetic logic gates[100] *in situ*. While naturally used for gene expression, riboswitches have been engineered and expressed *in vivo* to, among other functions, allow the real-time monitoring of cellular metabolites[101–103] and the modulation[104] or *de novo* engineering[105,106] of signaling pathways.

1.5.2 Fluorescence-Enhancing Aptamers

An important class of aptamers has been recently developed that binds to and activates the fluorescence of weakly- or non-fluorescent small molecule dyes. As mentioned in Section 1.1, these dyes consist of an electron donor and acceptor connected through a large, π-conjugated network. When these chromophores are unrestricted, the energy from absorbed photons is able to be released through non-radiative processes such as vibration or rotation. However, upon restriction of intramolecular rotation[18] by interaction with an aptamer, the absorbed energy can be made to release through radiative processes. The first fluorescence enhancing aptamer[107] was against malachite green (MG) and was found to enhance the fluorescence of MG by 2400-fold above background[55]. Since then,

fluorescence enhancing aptamers against thiazole orange[108] and derivatives of 4-(p-hydroxybenzylidene)imidazolidin-5-one (HBI)[109–111] - the active molecule in green fluorescent protein (GFP) - have been discovered and widely applied *in vivo*[101–103,112]. In much the same way that Green Fluorescent Protein(GFP) has allowed momentous discoveries in protein biology and cellular processes[10,113], these unique type of functional RNA promise to allow spatio-temporal detection and mapping of RNA *in vivo* and unlock a new era in nucleic acid biology and biotechnology.

1.6 Obstacles in Aptamer Development

Despite the discussed benefits of using nucleic acids and promises of amazing functionality, aptamers are usually relegated to use in only a few niche applications and have rarely found utilization outside of basic research. This lack of widespread acceptance is generally attributable to 1) the limited chemical functionality of nucleic acids and 2) the fact that conventional SELEX is time-consuming, labor-intensive and often fails to produce aptamers with desirable functionality[29,34].

1.6.1 Chemical Limitations of Nucleic Acids

Nucleic acids have inherently less chemical diversity than amino acids. For this reason, it is thought that natural nucleic acids may never be able to reliably address all possible targets and functions that proteins can. To confront this limitation, many groups have successfully incorporated non-natural nucleobases, backbones, sugars, and functional groups onto the nucleic acid scaffold in order to achieve enhanced affinity, specificity, function, and stability[60–63,114]. While this approach is a valid way of improving the chemical diversity of nucleic acids, there is a limit to the types and extents of modifications that can be made due to incompatibility of modified nucleotides with available biological machinery such as polymerases[60,63,115]. While it is possible to evolve new enzymes that accept these modifications[63,116], there does not exist a method of arbitrarily modifying aptamers and assuring that they are compatible with available enzymes. For many applications including *in vivo* uses[117], modified nucleic acids will likely remain necessary for widespread adoption so development of new and improved *in vitro* selection techniques that can readily incorporate new modifications remains a priority. While improving nucleic acid functionality with chemical and enzymatic modifications remains a topic of intense interest, this dissertation will not go into extraordinary detail of how to approach these challenges.

1.6.2 Inefficiencies in the SELEX Process

The SELEX process was a truly inspired discovery and has since allowed numerous insights into the mechanisms and roles of functional nucleic acids. However, it is not without issues.

Sequence Partitioning

Perhaps the most important step of SELEX is the efficient partitioning of functional from non-functional sequences (Figure 1.2 Step 2). Usually, this is performed by stringently washing away unbound sequences or those that do not adhere tightly with the ligand of interest. In this step, only aptamers which are tightly bound to the target survive and are chosen to undergo further rounds of selection. While simplistic in nature, this process is implicated as the cause for many of the problems that SELEX faces.

Under ideal equilibrium conditions, the likelihood that an aptamer remains bound to its target (and is thus recovered for further rounds of SELEX) is dependent on:

$$\ell_b = \frac{[T]}{[T] + K_d} \tag{1.1}$$

where ℓ_b is the likelihood that a described aptamer is bound to its cognate ligand, $[T]$ is the concentration of available target, and K_d is the dissociation constant of

the aptamer sequence with the target. Thus, every wash step in the partitioning process assists in removing high K_d (low affinity) sequences but also has a stochastic likelihood to remove the desired low K_d (high affinity) aptamers. For rare sequences, this is especially troubling because the prospect that an individual aptamer sequence remains is coupled to its representation (copy number) in a pool. Given that high-performing aptamers generally have many distinct regions necessary for their function[64], added functionality implies increased rarity in a random library. Thus, the rarest (and potentially most functional) sequences have a tangible chance of being irreversibly lost from the pool during this step.

In addition, SELEX methods reliant on equilibrium kinetics and partitioning in this manner all suffer from a theoretical limit on how strongly a well-performing sequence can be enriched relative to other, poorer-performing, sequences. Irvine and others have shown that the theoretical maximum enrichment for a given aptamer relative to another in the same round of SELEX, is equal to the ratio of their K_d[118–120]. Thus, successful enrichment of a high affinity aptamer is constantly being impeded by low-affinity sequences that are more common. This means that many rounds of SELEX are necessary to adequately enrich the highest affinity binders such that they can be identified during sequence analysis. Many rounds of selection inevitably lead to undesired biases in the process; such as the loss of rare, high-performing sequences[121], PCR bias[122], and parasitic amplification of

undesired sequences[118,119]. When compounded, these effects often lead to outright failure or identification of sequences with only poor performance.

While separation based on tight target binding has proved to be an effective way of finding aptamers for the purpose of target recognition (high affinity), more complex functions such as structure switching, fluorescence enhancement, target specificity, and catalysis are often identified only once the selection is completed[53,108,110]. Because the partitioning process does not take into account these higher-order functions, successful identification of functional aptamers is too often reliant on chance rather than experimental rigor.

To achieve these functions, sequences must often undergo further design and engineering after being discovered[123,124]. This can be done through incorporation of exogenous nucleobases[125]; or by introducing regions which induce structure-switching upon the addition of target ligands[104,123,126] or complementary oligonucleotides[53,92,127]. Optimization processes such as these mean that the aptamer discovery process does not end with sequence characterization and further work is needed to incorporate new functions[51,101,128]. Because added work begets added resources and costs, this process of rational design can be detrimental to the widespread application of aptamers. To avoid this issue entirely, there is growing interest in developing methods of directed evolution directly for aptamer function[129,130].

Sequence Analysis

Until recently, sequence analysis began with the identification of only a handful (typically less than 100) sequences using Sanger sequencing. These were then analyzed to identify conserved motifs indicative of their function. In comparison to Sanger sequencing, high-throughput sequencing (HTS) has increased the number of sequences which can be analyzed by several orders of magnitude ($> 10^6$ sequences) and is readily available to researchers through commercial vendors or academic facilities.

With the rise of HTS, it is becoming common practice to sequence every round's pool in a selection and analyze the resulting data using ever-more-functional software[131–133]. This is done in order to identify sequences that are 1) highly abundant, 2) exhibit a large increase in copy number over successive rounds or 3) predicted to function based on sequence-structure motifs because these sequences are most often the best-performing. Nonetheless, definitively identifying the singular best sequence from the hundreds or thousands of unique sequences resulting from HTS remains an unsolved issue.

Aptamer Characterization

Characterization is the bottleneck in many discovery processes because effective high-throughput characterization methods are not widely accessible. Thus, only a

handful of the sequences identified using HTS are usually subjected to rigorous characterization. An ideal, high-throughput characterization method would be done inexpensively (on a per-aptamer basis), using widely-available equipment and reagents, and with a minimal amount of solution-phase target. Previously, this had been done using custom DNA microarrays[134,135] on which thousands of aptamer sequences are printed into "colonies" in a uniform and defined manner. These chips can then be incubated with various concentrations of fluorophore-labeled target and the fluorescence of each colony can be correlated to binding performance. However, array-based methods are limited by the number of sequences that can be readily printed, and the length of DNA sequences that can be robustly synthesized due to inefficiencies of the phosphoramidite chemistry used in their production.

These limitations have been largely addressed through a high-throughput characterization method termed HiTS-FLIP[136]. This assay performs simultaneous aptamer discovery and characterization on millions of DNA sequences simultaneously using modified HTS equipment. While not yet available to most researchers through a commercial vendor or otherwise, this platform has since been expanded to work with RNA aptamers[137,138] and promises to be a very effective method for truly high-throughput RNA aptamer characterization. As an added benefit, characterization directly using the chip on which a pool is sequenced means that much of the guesswork required in identifying suitable candidates is removed.

1.7 Incorporating Design and Bias to Improve Directed Evolution

The limitations of SELEX are apparent when we compare the best *in vitro* selected aptamers with functional sequences that have been evolved *in natura*. While some RNA has been evolved to perform catalysis, their functionality is dwarfed by that of RNA in the ribosome (containing $\approx 3-7$kilo-bases of RNA). Due to its unfathomable theoretical complexity ($> 4^{3000}$) it is entirely implausible that the ribosome was developed *in loco*, rather it was likely assembled piece-wise through the random agglomeration of the many smaller, distinct catalytic RNAs that are believed to have once made up the RNA world[139].

Towards the goal of discovering novel higher-order nucleic acid function, we can mimic this natural piece-wise assembly. Indeed, there already exists an abundance of aptamers against relevant and interesting targets including small molecules, proteins, viruses, whole cells, tissues, and inorganic materials (See Section 1.3). However, these aptamers are oftentimes less functional than desired due to poor specificity, inadequate affinity, or lackluster performance[33]. While this problem has lead to an abundance of new and better selection methodologies[64], it is rarely

desirable to evolve a new aptamer or design a discovery method every time a new application is explored.

Instead, a known aptamer can be modified to yield new or optimized performance.This is usually done through two methods: 1) rational modification of existing function from known aptamer motifs, or 2) biasing an aptamer pool to explore nearby sequence space (i.e. a sequence/performance landscape).

Rational modification of known motifs can be used to explore function of an aptamer by point- or segment-wise modification[136–138,140,141]. For example, a natural guanine-specific riboswitch was found to switch specificity to adenine on the basis of a single C to U substitution[142]. In general, the complexity of any folding-binding relationship for an arbitrary nucleic acid currently exceeds our rapid and accurate computational and design capabilities[143–145]. Thus, for any given function, designing an optimized functional motif from first principles is extremely difficult - and only becoming a more distant goal as we aim for higher-order complexity.

Biasing a pool is usually done through pool mutagenesis[146,147] using methods such as error-prone PCR[148] or random splicing/recombination of sequences in a pool[149,150]. However, design principles can play roles in designing new random libraries[151] for *in vitro* selection. For example, a library incorporating minimally-conserved structural regions resulted in better-performing aptamers than when

using a totally random library[152]. Thus, in many cases, it could be ideal to develop a partially-designed random library around a known target aptamer and perform a selection using the best available methodology so as to improve the aptamer and incorporate additional function. However, this requires we move beyond the inefficiencies of traditional SELEX.

1.8 Particle-based Display Methods To Overcome SELEX Limitations

In 2014, Wang et. al presented a method termed particle display (PD) that promises to revolutionize the discovery of aptamers[153]. Using this method, they were able to reliably obtain aptamers against targets previously believed to be untargetable using natural DNA aptamers[60], and with affinities rivaling or exceeding those of the best available antibodies. The reason for this methods incredible success is because of how it addresses many of the problems that riddle traditional SELEX.

1.8.1 The Particle Display Process

Particle display begins by transforming a solution-phase pool of random DNA sequences into homogeneously-coated, monodisperse aptamer particles (APs, Figure 1.4). These APs display thousands of copies of a single DNA aptamer on the

surface of paramagnetic, 1μm polystyrene beads. These are then incubated with a solution-phase, fluorescently-labeled protein. Next, each AP is interrogated using a Fluorescence-Activated Cell Sorter (FACS) which measures fluorescent and light-scattering properties of individual particles and can separate them based on parameters defined by their fluorescence signature. Thus, APs which exhibit high fluorescence (indicative of target recognition) can be quickly and effectively separated from non-fluorescent APs. Collected particles are then subjected to amplification via PCR and the cycle repeats. Whereas traditional SELEX usually requires 6-20+ rounds of selection, PD can reliably produce high-quality aptamers in 3 rounds or fewer. This eliminates many of the biases associated with repeated rounds of SELEX, and greatly improves the success rate of aptamer discovery.

1.8.2 Advantages of Particle Display

The reason that PD can reproducibly achieve aptamers with remarkable performance is because it almost entirely removes the theoretical limit on enrichment possible for a well-performing aptamer over its poorer-performing counterparts during a single round of partitioning (as discussed on page 20). Whereas the maximum enrichment of one sequence relative to another in SELEX is tied to the ratio of their K_ds, PD is capable of enriching sequences at levels that are orders of magnitude more efficient (Figure 1.5). This is because each aptamer is present in

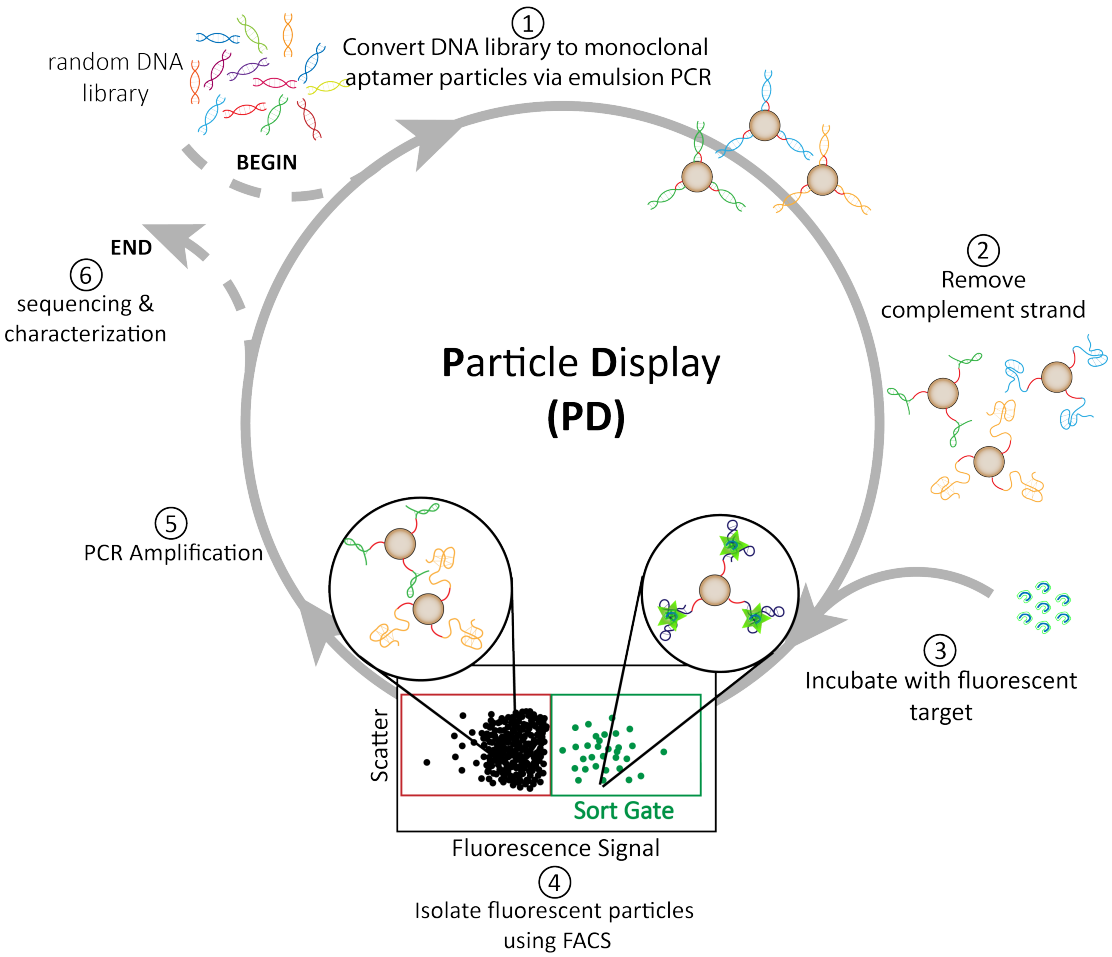

Figure 1.4: Selection scheme using Particle Display (PD). **Step 1**: A random pool of DNA is converted into monoclonal aptamer particles (APs) using emulsion PCR. **Step 2**: Emulsions are broken, excess reagents are washed away, and the complementary DNA strand is removed. **Step 3**: APs are incubated with a fluorescently-labeled target **Step 4**: FACS is used to isolate fluorescent APs from those with little or no fluorescent signal. **Step 5**: Collected APs directly undergo PCR amplification. Steps 1-5 are repeated 2-3 times until pool converges to a handful of functional sequences. **Step 6**: Pool undergoes sequencing and candidate aptamers are characterized.

huge populations that are individually characterized before being partitioned. In this step, only functional populations that exhibit high target binding (indicated by fluorescence response) are collected for the next round of PD. This is a remarkable improvement over SELEX wherein the vast majority of retained sequences in the early rounds of selection are weakly- or entirely non-functional.

PD also removes the uncertainty associated with equilibrium binding kinetics common in traditional SELEX. When an aptamer is present in a single copy in solution, the likelihood that it is bound is dependent on Equation 1.1. Thus, even high-performing (low K_d) sequences have a non-zero chance of being irreversibly lost from the selection. In PD, however, ℓ_b defines not the likelihood that an aptamer is recovered during a round, but rather the fraction of aptamers displayed on an AP that are bound to their target:

$$f_b = N \cdot l_b = \frac{N \cdot [T]}{[T] + K_d} \tag{1.2}$$

where N is the number of sequences expressed on the surface of an AP and is on the order of $\approx 10^5$ per AP. When an AP is measured using FACS, the mean fluorescence signal is defined as:

$$F = (F_{max} - F_{bg}) \cdot f_b + F_{bg} = (F_{max} - F_{bg}) \cdot \frac{N \cdot [T]}{[T] + K_d} + F_{bg} \tag{1.3}$$

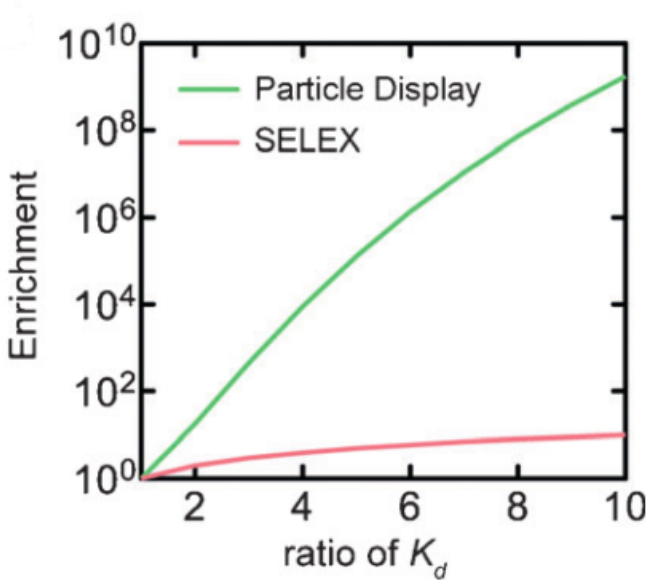

Figure 1.5: Maximum enrichment ratio possible for two sequences as a function of the ratio of their K_d. PD exhibits orders of magnitude higher enrichment potential, explaining how it is able to converge a pool to a small number of highly functional sequences over the course of 2-3 rounds. Reproduced with permissions[119].

where F_{max} is the maximum mean fluorescence of a particle saturated with target, and F_{bg} is the mean background fluorescence due to non-specific target binding.

In PD, only particles which exhibit a fluorescence signal F greater than the sorting threshold defined as $F_{max}/3$ are collected. If a significant fraction (typically considered as greater than 0.1%) of the APs surpass the sorting threshold, the target concentration is reduced (Figure 1.4. Step 3) such that fewer APs fluoresce at levels above background. Thus, rare, well-performing sequences can be greatly enriched even in the presence of more abundant, low- or moderately-performing sequences by simply adjusting $[T]$ such that the fluorescent signal for these undesirable sequences falls below the sorting threshold.

Using this method, Wang et al. discovered DNA aptamers again 4 distinct protein targets with K_d values ranging from 7pM to 2.3 nM[119]. Interestingly, three of these targets had been previously classified as "untargetable" using natural DNA aptamers[60].

In addition to these remarkable abilities, PD has been further modified to allow selection of highly-specific DNA aptamers against protein targets. In 2017, Wang et al. showed that by enriching for sequences in the presence of fluorescently-labeled non-target proteins, aptamers can be discovered that show unrivaled specificity to target proteins[154]. These sequences show better affinity and specificity in complex environments (e.g. 10% human serum)than even the best commercially-available

antibodies, and shows that PD-based methods of aptamer discovery can be used for the direct, functional selection of incredibly well-performing aptamers.

1.8.3 Limitations of Particle Display

While particle display has exhibited some remarkable abilities to discover high-performing aptamers, it suffers from two major drawbacks. First, PD is currently only accessible for DNA-based selections. While there exists much debate over the advantages and limitations of DNA[76], it is generally believed that RNA can exhibit higher-order functionality due to the 2′-OH group and the more diverse structures into which it can fold[155]. Additionally, RNA aptamers have direct relevance *in vivo* because they can be expressed directly by cellular machinery whereas single-stranded DNA aptamers cannot.

Secondly, the throughput of APs separable by FACS is on the order of 2×10^7 per hour. Thus, the maximum number of sequences which can be exhaustively measured in the first round of PD is $\approx 10^8$. While this number could be improved by 1-2 orders of magnitude through parallelization efforts over long time scales, it is dwarfed by the number of sequences typically searched using traditional SELEX ($\approx 10^{14}$ sequences). As we look to push the limits of functionality and find aptamers exhibiting ever-more-complex activity, it will become necessary to perform PD-based selections using highly diverse libraries. Currently, this is

addressed by performing a small number of SELEX rounds (typically 1-2 rounds) prior to initializing PD. This helps reduce the diversity of a library to a number more manageable by FACS without suffering the drawbacks associated with many rounds of SELEX (see section 1.6). However, many issues remain with traditional SELEX that lead us to prefer an exclusively particle-based selection technique.

1.9 Thesis Objectives

Aptamers have shown incredible functionality since their initial discovery in 1990. However, many problems remain with traditional SELEX methodologies that make them incompatible with desired functional selections or inefficient at reliably discovering new aptamers against the plethora of interesting targets. As we begin to require more complex and functional nucleic acids (such as fluorescence enhancement or switching), directed evolution using traditional SELEX will continue to lag behind our needs. Particle-display based methods have been shown to answer many of the issues that SELEX encounters, but comes with its own drawbacks including limited library size and lack of functional diversity. In this dissertation, I summarize a series of projects towards addressing these remaining issues.

I explore the uses of a new, particle-based selection technique, termed GRAP display, towards efficient discovery of functional RNA aptamers. Chapter 2 introduces this new technique and shows how it can be used to discover highly-functional,

unique, and applicable fluorescence-enhancing aptamers against the small molecule dye malachite green.

Chapter 3 describes a method of addressing throughput limitations of particle-based selection methods through incorporation of a known functional aptamer motif. I show that, without the use of any traditional SELEX methodologies, I can discover aptamers that outperform the best known fluorescence-enhancing aptamer in terms of affinity and fluorescence enhancement. I also introduce a new method of random library design that, coupled with GRAP display, could be applied to maximally explore sequence space around known functional sequences.

Chapter 4 combines the topics put forth in Chapters 2 and 3 to show that GRAP display can be used to discover structure-switching RNA aptamers through the re-evolution of known motifs joined by a randomized transduction region. The results of this selection show that GRAP display can discover switching motifs beyond those that would be predicted using rational design. This selection is the first step towards evolving modular aptamers of complex and arbitrary function.

Chapter 5 summarizes these results and briefly introduces possible directions for further research towards functional and efficient aptamer selection.

Chapter 2

Gene-linked RNA Aptamer Particle (GRAP) Display

2.1 Introduction

Fluorescent proteins have an indispensable role in any biologists toolkit. This is because these proteins allow an unparalleled view into the expression, localization, and activation of cellular processes. Green fluorescence protein (GFP)[156] is arguably the most famous fluorescent protein and has resulted in countless discoveries, earning the founders of this protein the Nobel Prize in Chemistry in 2008[113]. While a vast array of fluorescent proteins are available to researchers

today[157,158], there exist very few methods of fluorescently-labeling a biomolecule of equal importance: RNA.

GFP works through encapsulation of 4-(p-hydroxybenzylidene)imidazolidin-5-one (HBI)[156], which is formed spontaneously from the tripeptide Ser65-Tyr66-Gly67[159]. In general, fluorescent proteins work through activation of an internalized, covalently-linked small molecule chromophore. (i.e. HBI) These chromophores are non-fluorescent but become active upon correct folding of their protein scaffold[160]. When excited with light at one wavelength, the encapsulated chromophores fluoresce and emit light at a lower energy, red-shifted wavelength[16].

Because no sequence of natural RNA nucleotides is known to naturally fluoresce, fluorescent RNA requires the addition and subsequent activation of an external chromophore[13]. Thus, when we wish to image an RNA of interest, it is oftentimes fused with a fluorescence-enhancing (FE) RNA aptamer that activates an externally-applied chromophore. The target RNA can then be localized by detection of the FE RNA aptamer. The Spinach aptamer[109], for example, enhances the fluorescence of DFHBI (Figure 1.1) and has been used to image small-molecule metabolites[101–103,161] and mRNA expression[112]; among other uses.

However, unlike fluorescent proteins, only a handful of FE RNA aptamers are currently available[13]. This puts significant limitations on their utility in biological research. This problem is a direct result of inefficiencies in the selection procedure,

because SELEX enriches aptamers foremost on their ability to bind a target. It is only after a selection is performed that these sequences can be tested for their FE ability. Thus, the discovery process is inherently inefficient and low-yield. For example, less than 1% of the aptamers discovered alongside RNA Spinach produced any measurable fluorescence signal[109,110]. To compound issues, previous selections have shown that the binding affinity of an aptamer correlates poorly to the aptamers fluorescence enhancement capabilities[108,109]. For these reasons, there is an unmet need for a selection method that can reliably and efficiently produce fluorescence-enhancing aptamers based foremost on this ability. Successful deployment of this method would enable accelerated discovery in nucleic acid biology in a similar manner to what the widespread availability of fluorescent proteins has done for protein biology[10,113].

Our lab has previously demonstrated the utility of particle-based display methods for the selection of DNA aptamers using particle display (PD) against fluorescently-labeled protein targets[119,154]. PD offers many advantages, such as the ability to interrogate each aptamer sequence individually and efficiently separate only those sequences with the desired functionality. Building off this work, we wished to develop a particle-based display method for the expression of RNA aptamers. We hoped to develop this method in order to take advantage of the

many benefits of particle display (see Section 1.8.2), and to prove its usefulness by selecting for a function which traditional SELEX has difficulty addressing.

To this end, we have developed **G**ene-Linked **R**NA **A**ptamer **P**article (GRAP) display to enable the screening of RNA aptamers directly for their ability to cause FE of a small molecule chromophore - malachite green (MG). This is the first *in vitro* selection methodology that allows high-throughput screening of RNA aptamers based upon fluorescence performance rather than binding affinity. We use this method to identify aptamers with higher affinity and quantum yield than the best known fluorescence-enhancing aptamer. In addition, we show that this method allows us to tune the emission profile of the selected aptamers, a first step in being able to offer a palette of *well-performing*, fluorescence-enhancing (FE) RNA suitable for biologists use.

2.2 Results and Discussion

2.2.1 Gene-Linked RNA Aptamer Particle (GRAP) Display Process

In PD, 1 μm paramagnetic polystyrene beads are modified with a DNA primer sequence to allow the display of covalently-linked, single-stranded DNA sequences. In GRAP display (Figure 2.1), we modify these same 1 μm beads to incorporate an

additional oligonucleotide which allows us to tether an RNA sequence to the bead surface. Specifically, magnetic beads are functionalized with equivalent amounts of a DNA forward primer and a poly-thymidine (poly-T) capture oligo which is terminated with a 3′ inverted deoxythymidine to inhibit extension during PCR. These beads then undergo an emulsion PCR process whereby the particles are converted into monoclonal gene particles that each display thousands of identical copies of a fluorescently-labeled DNA sequence on their surface. The DNA on these particles remains double-stranded, and the beads are subjected to an additional emulsion reaction under *in vitro* transcription conditions to yield a pool of GRAPs (see Methods).

Upon breaking these emulsions and removing reaction byproducts, these particles are incubated with the non-fluorescent dye and a fluorescent signal for each particle is obtained through FACS. If the fluorescence signal of a particle surpasses the background fluorescence as described in Section 1.8.2, the FACS instrument physically separates this particle from its non-fluorescent counterparts. Only these collected particles are added to a polymerase chain reaction (PCR) mixture in order to amplify the displayed DNA. Amplified DNA product is then used for further rounds of GRAP display until no more enrichment is observed. Because we are screening directly based on the fluorescent signal of an aptamer-chromophore

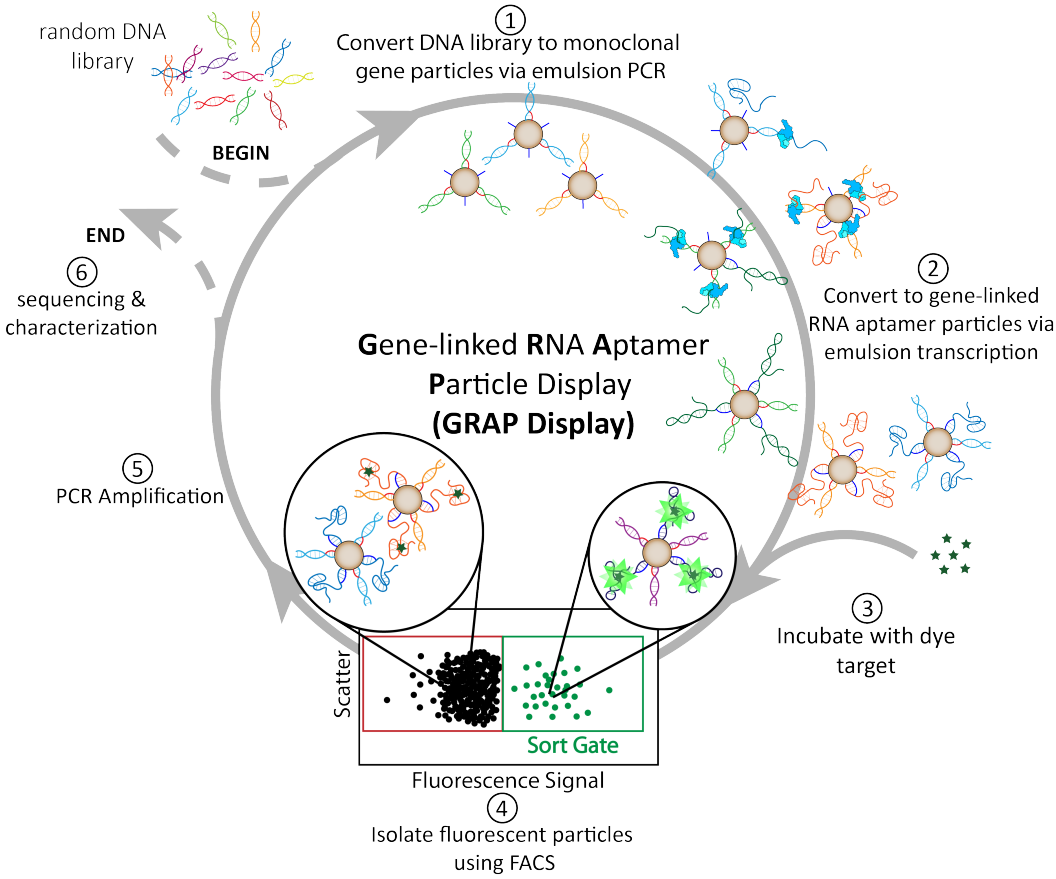

Figure 2.1: Gene-linked RNA Aptamer Particle (GRAP) display. **Step 1**: A random pool of DNA is converted into monoclonal gene particles using emulsion PCR. **Step 2**: Emulsions are broken, excess reagents are washed away, and the particles are re-emulsified with transcription reagents. **Step 3**: Emulsions are cleaned and excess reagents removed. Gene-linked RNA Aptamer Particles (GRAPs) are incubated with a non-fluorescent chromophore **Step 4**: FACS is used to separate GRAPs with high fluorescence from those with no fluorescence. Particles that do not bind or bind but do not enhance chromophore fluorescence are not collected. **Step 5**: Collected GRAPs undergo PCR amplification. Steps 1-5 are repeated 2-3 times until pool converges to a handful of functional sequences. **Step 6**: Pool undergoes sequencing and candidate aptamers are characterized.

complex, this method is effective at screening directly for FE aptamers where other selection methodologies fail.

Enrichment is monitored by looking at the fraction of DNA-coated GRAPs that surpass background fluorescence at a given target concentration. For subsequent rounds, if there is no marked increase in the signal of beads surpassing the sorting threshold (Figure 2.1, Step 4), the final pool is deemed converged. If, however, more GRAPs surpass this threshold than in the previous round, an additional round of screening is performed at a lower target concentration in order to increase selection stringency. Once no further enrichment is observed over subsequent rounds (which usually takes 2-3 rounds using GRAP display), all pools in the selection undergo HTS.

2.2.2 GRAP Display Allows Fluorescence Characterization of Individual RNA Aptamers

In contrast to conventional SELEX-style methods that enrich aptamers solely based on binding, GRAP display allows us to profile and isolate aptamers directly based on their fluorescent properties. The signal (F_λ) generated by a given aptamer upon binding malachite green is determined by several aptamer properties.

For example, a high F_λ at a measured wavelength may arise because the aptamer has a high quantum yield (high ϕ_F), is highly occupied (low K_d), or because the

excitation and emission profile results in many photons emitted at the measured wavelength (high S_λ) upon excitation. Likewise, a low F_λ could result from low quantum yield, poor affinity, or because the excitation or emission spectra falls outside the excitation wavelength or emission channel. The relationship between these various properties is described by the equation:

$$F_\lambda = \phi_F \cdot \frac{[T]}{[T] + K_d} \cdot S_\lambda \tag{2.1}$$

where [T] is the concentration of malachite green.

In SELEX-style experiments, aptamers with high affinity (low K_d) for the target molecule are preferentially enriched. This can lead to a selection failing to yield suitably bright aptamers because mere binding to an aptamer is insufficient to guarantee restriction of intramolecular rotation (RIR). By analyzing each aptamer based on the properties described, we show that GRAP display can be used to tune selections for fluorescence-enhancing aptamers based on their unique combination of fluorescent properties.

2.2.3 Rationale for Choosing Malachite Green

We performed our first GRAP display selection using malachite green (MG). We chose MG as our target for four reasons: i) MG undergoes fluorescence enhancement

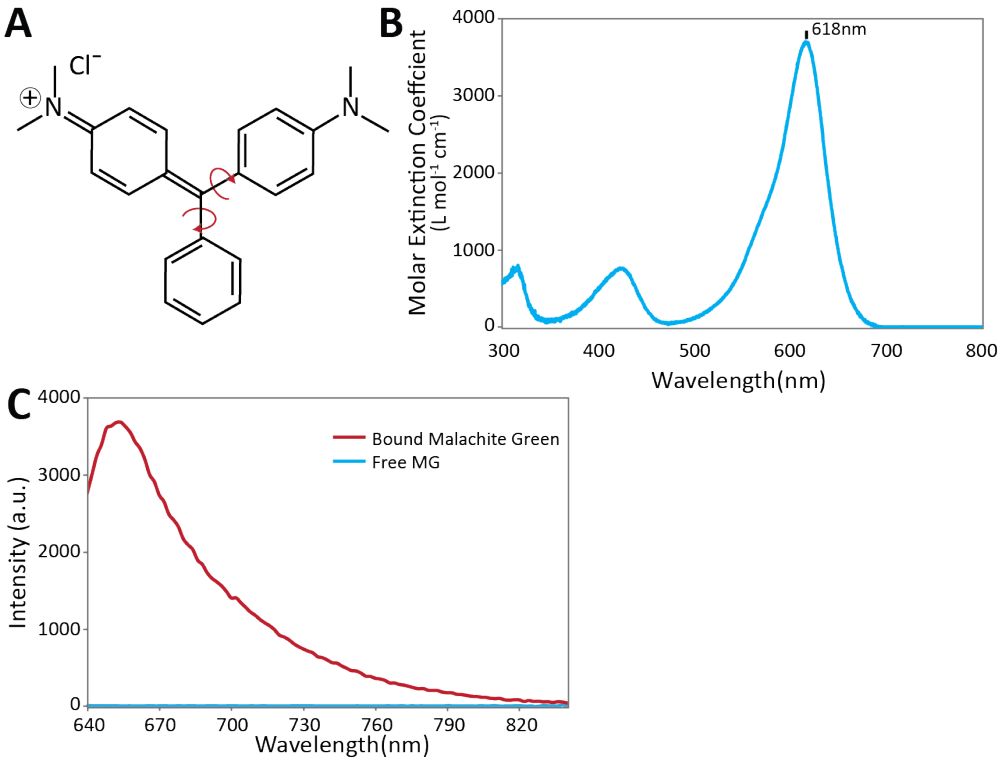

Figure 2.2: A) Chemical structure of malachite green (MG). When unconstrained, the aromatic rings are able to rotate freely (red arrows). **B)** Absorbance spectra of free MG. MG exhibits a maximum absorbance peak around 618nm ($\epsilon = 148,900 M^{-1} cm^{-1}$)[162]. **C)** Fluorescence emission spectra of aptamer-bound MG. Upon binding to **MGA**, the fluorescence of the MG:MGA complex (red) increases by about 2360-fold above MG in the unbound state (blue)

upon RIR (Figure 2.2 A), ii) The maximum absorbance band[152] is compatible with the FACS laser used at 633nm (Figure 2.2 B), iii) MG in the unbound state is very weakly fluorescence ($S_{\lambda}^{unbound}$) (Figure 2.2 C (blue spectra)), and iv) the brightest-known fluorescence enhancing aptamer is against MG[55,107](Figure 2.2 C (red spectra)). Thus, we wished to challenge our methods ability to find better (brighter, higher affinity) aptamers than SELEX allows.

2.2.4 GRAP Display Enriches for FE Aptamers with Unique Fluorescent Properties

We began the GRAP display process by first performing two rounds of traditional SELEX against MG in order to reduce the diversity in our pool to a size more manageable by FACS ($\approx 10^8$ sequences). While SELEX has many disadvantages and enriches based upon binding affinity rather than fluorescence enhancement, many studies have shown that the early rounds of SELEX are efficient at enriching towards functional aptamers without suffering from the biases that many rounds of SELEX incorporates[132,163,164].

Prior to initializing the GRAP display process, we were aware that fluorophores can adopt unique excitation and emission profiles dependent on their micro-environment[16]. Also, previous selections have shown that a single chromophore can produce distinct fluorescence profiles based on the aptamer sequence with which it is bound[109]. Thus, we attempted to actively select for aptamers that produced distinct emission profiles.

The partially-converged pool was converted into GRAPs through two consecutive emulsion reactions. The resulting GRAPs express a functional RNA aptamer and their fluorophore-labeled parent DNA. Next, we incubated these with 10 μm MG and collected all APs which exhibited fluorescence above background at

wavelengths greater than MG's primary excitation peak of 618nm (Figure 2.2 B). Specifically, this entailed collecting particles from three sort gates: 670 ± 30nm, 730 ± 45, and 780 ± 60nm. Particles collected in these gates were combined and underwent PCR in preparation for the next round of GRAP display.

Over the course of selection, the concentration of malachite green was reduced from 10 μM at the start (Round 0), to 1 μM in Round 1, and 100 nM in Round 2 (Figure 2.3). The mean fluorescence intensity above background of the resulting Round 3 GRAPs was dramatically improved over the Round 0 pool, as indicated by the significantly higher fraction of particles meeting the sort gate threshold even at a drastically lower target concentration ($7.73 - 14.3\%$ in Round 3 compared to $0.03 - 0.13\%$ in Round 1). With the Round 3 GRAPs, no additional enrichment was observed at concentrations below 100 nM.

In addition to greatly enriching for fluorescence-enhancing aptamers, we noticed that the Round 3 pool had bifurcated into two distinct populations with separate emission profiles. We partitioned the Round 3 pool into a blue-shifted (Round 3-B) pool and a red-shifted (Round 3-R) pool based upon the relative intensities in the 670 ± 30nm and $780 \pm 60nm$ channels(Figure 2.4).

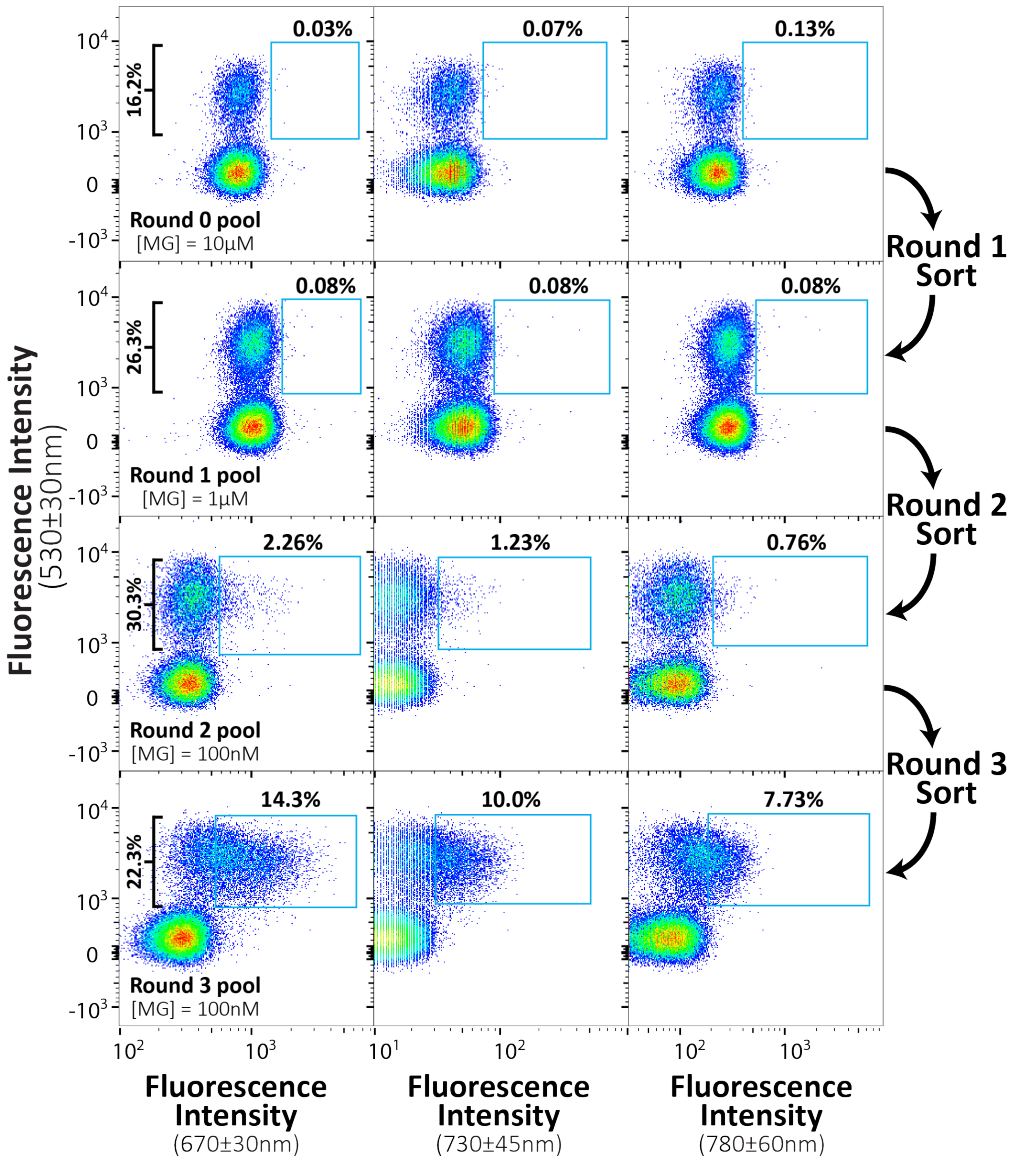

Figure 2.3: FACS summary of GRAP display process. We performed three rounds of GRAP display and collected all events that exhibited a strong DNA-fluorophore signal (y-axis) and fluoresced above background in the $670 \pm 30nm$, $730 \pm 45nm$, and $780 \pm 60nm$ emission channels (x-axis). Particles sorted in each channel were pooled before PCR amplification and the next round of GRAP display. We observed no further enrichment in the Round 3 pool compared to the Round 2 pool at concentrations below 100 nM malachite green. As target concentration is reduced, the sort gate is shifted to the left in order to account for reduced background fluorescence.

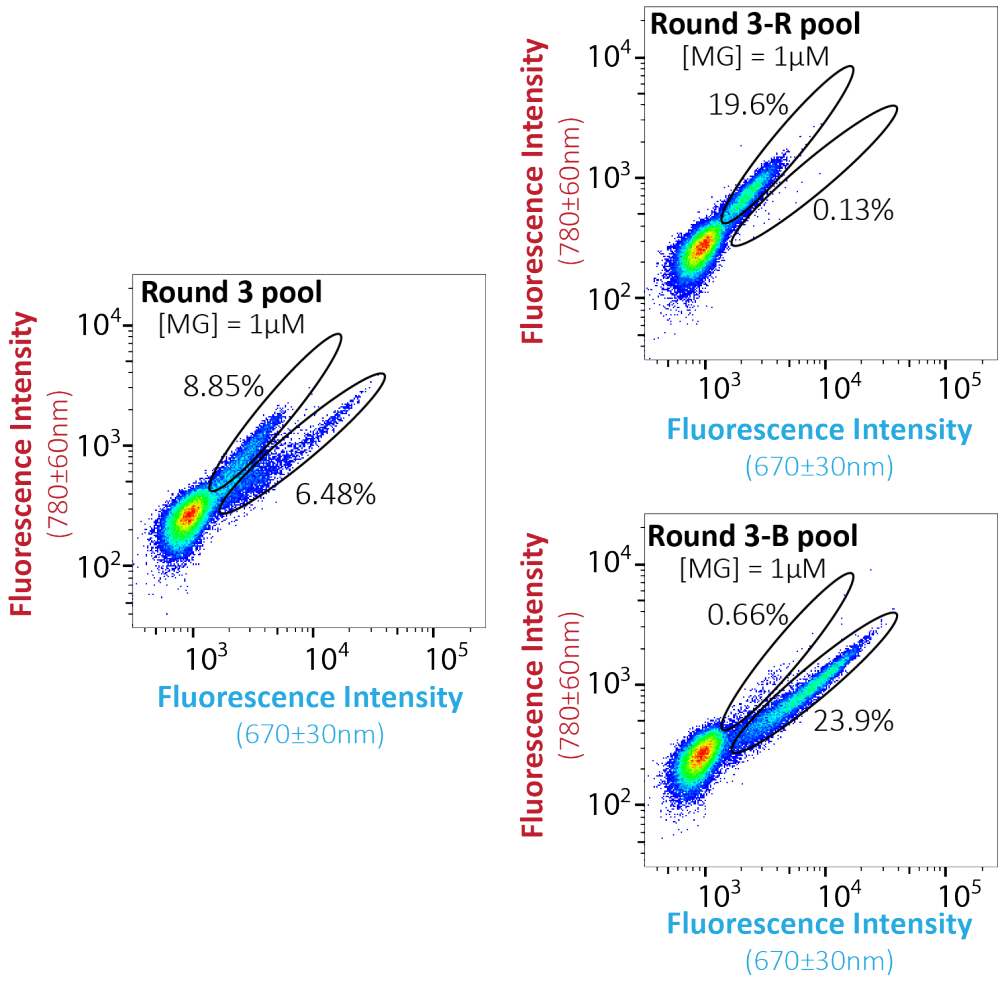

Figure 2.4: FACS plots of the Round 3 GRAP pool. This pool was clearly bifurcated into two populations exhibiting distinct emission profiles. Upon separating with FACS, one pool (Round 3-B) exhibited a higher mean fluorescence in the $670 \pm 30nm$ channel, whereas the other pool (Round 3-R) exhibited a higher mean fluorescence in the $780 \pm 60nm$ channel.

2.2.5 High-Throughput Sequencing Identifies Numerous FE Aptamers

We next subjected each round to HTS and obtained $2.44.2x10^6$ sequences from the Round 2, Round 3, Round 3-B, and Round 3-R pools. To begin narrowing these candidates to a number more suitable for detailed characterization, we counted and recorded a reads-per-million (RPM) value for all sequences in each pool and chose to characterize the union of the top five most represented sequences in the Round 3, Round 3-B, and Round 3-R pools. This comprised 5 sequences from Round 3, none from Round 3-B (they were identical to the five most abundance Round 3 sequences already chosen), and 3 sequences from Round 3-R (See Methods for further explanation).

Next, we determined a fitness (i.e. enrichment: ratio of a sequences current round RPM over its previous round RPM) value for all sequences that were significantly represented in the raw sequencing data (greater than 20 copies). We determined this value because previous studies have repeatedly shown that high enrichment over a round of selection can be a better predictor of performance than copy number alone[132,163,164].

Many high-fitness sequences in each round were self-similar except by a one- or two-base point mutation, insertion, or deletion. For example, sequence **R3-B-3**

differs from sequence **R3-4** by a single point mutation. Relatedly, 95 of the top 100 highest-fitness sequences in the Round 3-B pool are highly sequence-similar to **R3-4** in the same manner. For this reason, we chose the top 5 (if available) non-similar sequences from the Round 3, Round 3-B, and Round 3-R pools. In total, this comprised 5 sequences from Round 3, 4 sequences from Round 3-B, and 5 sequences from Round 3-R. In total, 22 unique sequences were identified for further characterization (Table 2.1. See Table 2.4 in Methods for copy number and fitness information).

2.2.6 Fluorescence Performance of Identified Aptamers

In recording the performance of these sequences, we wished to decouple RNA concentration from binding affinity because a single RNA sequence can exist in multiple distinct thermodynamically-stable structures[155,165]. Thus, the concentration of RNA is not always equal to the concentration of correctly-folded (active) aptamer. For example, high concentrations of RNA can promote dimerization of two otherwise incompatible RNA strands by promoting more interaction between nearby strands.

For this reason, we set the concentration of RNA at 100 nM and performed a titration of free malachite green from 10 nM to 10 μM. The fluorescence response at each concentration was measured using a Tecan M1000 plate reader

Name	**Sequence** ($5' \rightarrow 3'$) of core region	**Effect of Poly-A Tail**
R3-1	TGCACGGAGGCGCCTCTGGTCCTTAGTTTAATCCATAACG	+
R3-2	TACAATCTCAGGTGCACCTTTAGACCACTGTGTCTCTACA	o
R3-3	CTAGCTTCGAGTAACGAAACGCGTGAGGCTGCCCCGCTCC	o
R3-4	AGTCTAGAGCGGGTACCTGTGCTGCAAATCTCCCGTCACT	+
R3-5	GTCCCAAGGAGGTGGGTGGGTGGTGTCGACTCCTTTTCTT	o
R3-6	GTGCCACCAGACTCACGCACTAGATCCGGCCGAGTAACGA	-
R3-7	GTCCCAAGGAGGTGGGTGGGTGGTGTCGACTCCTTTTCTT	-
R3-8	AGAAATTGTGTAAGACCTTATTGAATGAGGCGCAACTCGC	-
R3-9	TACAATCTTAGGTGCACCTTTAGACCACTGTGTCTCTACA	-
R3-10	CTAGCTCCGAGTAACGAAACGCGTGAGGCTGCCCCGCTCC	-
R3-B-1	TTACAATACAAGTCTTCGAAGACTGAGTTTTCCATCTCCA	+
R3-B-2	TTCATCAACATAGGAGTGGGAGGGGTAGCTAGGGCTCGTG	+
R3-B-3	AGTCTAGAGCGGGTACCTGTGCTGCAAATCTCCCGTCATT	-
R3-B-4	CCGGCGTACTACGCGTAACCATGATCCCCACTCTTGACTT	-
R3-R-1	ATCGATTCAATGCCGCCAAGGGCTAGCCCGATCCATAACG	o
R3-R-2	CGCGTGCAACAGGGCCCTGATCCAGCACTCACTTATAACG	+
R3-R-3	TACAAGAGATTCAAGCGCACGGTCTTCCACACCCAACTCG	+
R3-R-4	TGCACGGAGGGGCCTCTGGTCCTTAGTTTAATCCATAACG	-
R3-R-5	ATCGATTCAATGCCGCCAAGGGCTGGCCCGATCCATAACG	o
R3-R-6	CGCGCGGGTCTCGAGACCCTGAATCATTTCTCGAAAAACG	+
R3-R-7	CGCGTCACCAGGGACACTCGGATCCACTCACATCATAACG	o
R3-R-8	CGCGTGCAACAGGGCCTTGATCCAGCACTCACTTATAACG	+

Table 2.1: 22 sequences were identified from HTS for further characterization. The effect of the poly-A tail on each sequence was evaluated ('+' means the poly-A tail improves function, 'o' means it has no effect on function, and '-' means it has a negative effect on the aptamers function). Sequences were characterized without a poly-A tail unless it was found to improve function. See Methods Section 2.4.14 for further details. Sequence **R3-7** differs from **R3-5** by an A8' deletion in the FP region and was characterized as such.

and fit to Equation 2.1 (see Methods Section 2.4.8). For each sequence, we were able to obtain a dissociation constant (K_d) and signal generated by the bound chromophore:aptamer complex (F_λ) at saturation ($\ell_b \approx 1$).

We next calculated S_λ by integrating the corrected excitation and emission spectra for each aptamer relative to **MGA**. We could then compare each aptamer to **MGA**($\phi_F^{MGA} = 0.187$[55], 2360-fold fluorescence enhancement, $S_\lambda^{MGA} = 1$. See Methods for sequence information) and obtain quantum yields from Equation 2.1:

$$\frac{F(\lambda)_{sat}}{F(\lambda)_{sat}^{MGA}} = \frac{\phi_F}{\phi_F^{MGA}} \cdot \frac{S_\lambda}{S_\lambda^{MGA}} \tag{2.2}$$

Rearranging:

$$\phi_F = \phi_F^{MGA} \cdot \frac{S_\lambda^{MGA}}{S_\lambda} \cdot \frac{F(\lambda)_{sat}}{F(\lambda)_{sat}^{MGA}} \tag{2.3}$$

The results are shown in Table 2.2.

Of the 22 sequences identified by HTS, only **R3-3** exhibited no fluorescence. Two sequences, **R3-4** and **R3-R-3** enhanced the fluorescence of MG, but perform above the concentration range probed in our characterization. Despite these three sequences, this is a dramatic improvement over previous selections where as little as 1% of characterized sequences exhibited any measurable enhancement[110].

Interestingly, **R3-3** differs from **R3-10** by a single nucleotide point mutation (Figure 2.5) yet has no measurable K_d or FE. This suggests either 1) the point

Name	S_λ	K_d (nM)	**Fluorescence Enhancement** (fold increase)
R3-1	1.05	2017.9 ± 167.6	110.7
R3-2	0.88	42.7 ± 12.2	19.2
R3-3	n.d	n.d	n.d
R3-4	n.d	n.d	n.d
R3-5	0.87	92.6 ± 22.4	49.7
R3-6	1.17	111.6 ± 12.0	259.4
R3-7	0.87	125.8 ± 34.6	42.3
R3-8	0.98	29.3 ± 8.4	17.7
R3-9	0.88	38.6 ± 10.7	37.9
R3-10	1.13	220.4 ± 39.2	317.5
R3-B-1	1.11	320.8 ± 26.7	619.5
R3-B-2	1.11	144.9 ± 26.3	494.9
R3-B-3	1.24	2094.6 ± 347.3	219.2
R3-B-4	1.15	579.3 ± 75.5	4273.3
R3-R-1	1.05	1301.9 ± 345.1	275.1
R3-R-2	1.05	1549.1 ± 441.8	238.2
R3-R-3	n.d	n.d	n.d
R3-R-4	1.05	1045.2 ± 124.6	189.2
R3-R-5	1.05	1488.7 ± 370.4	78.7
R3-R-6	1.05	1827.3 ± 175.6	147.9
R3-R-7	1.05	1561.7 ± 221.5	198.4
R3-R-8	1.05	1590.7 ± 227.0	237.3
MGA	1.00	61.0 ± 13.4	2360.0

Table 2.2: Characterization results of identified sequences. An area scaling constant (S(λ)) was determined by integrating the corrected excitation and emission spectra of each sequence, relative to **MGA**. Equations 2.1 and 2.3 were used to calculate the K_d and fluorescence enhancement of each aptamer over background. These values assume perfect folding of all sequences, so may differ depending on the manner of determining K_d.

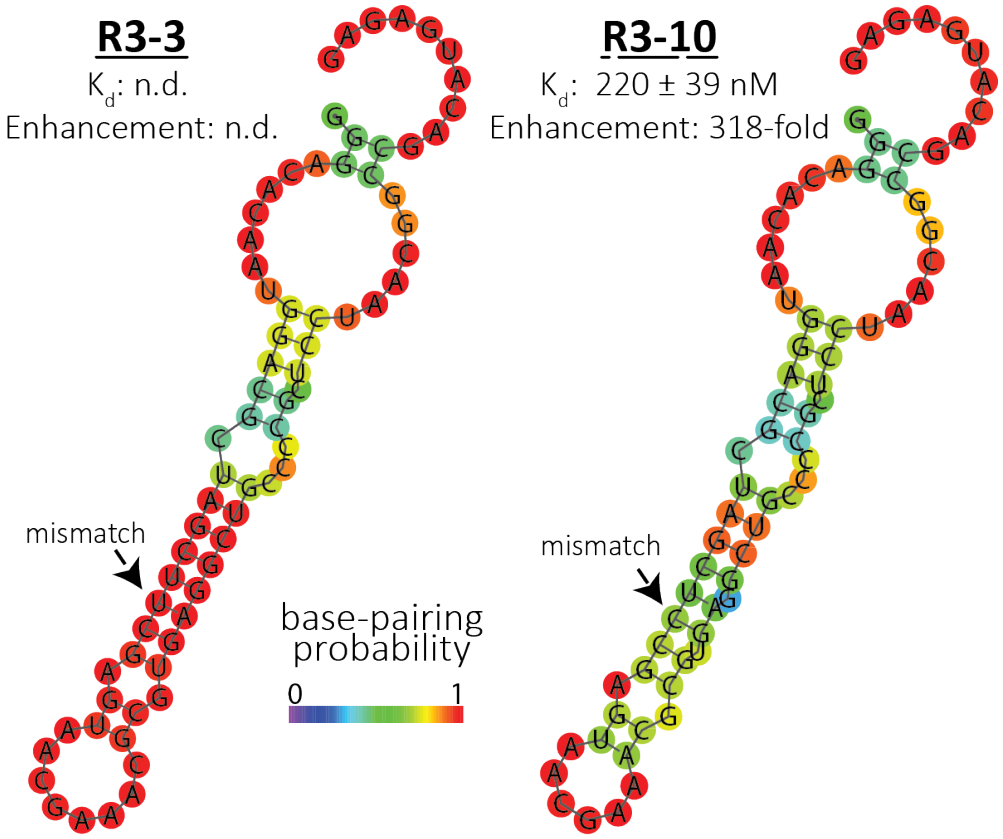

Figure 2.5: Predicted secondary structure of sequences **R3-3** and **R3-10** using mfold software[166,167]. These sequences differ from each other by a single mismatch, yet the activity of **R3-3** is not detectable (n.d.) suggesting the binding pocket is inactive. We hypothesize this to be a result of the C→U mutation causing base slippage and resulting in a more stable structure that disrupts the binding pocket for this aptamer.

mutation destabilizes the binding pocket, or 2) the nucleotide is responsible for interacting with MG. Based on the predicted mfold [166,167] structures, this mutation results in a significant change in structure. Thus, it seems likely that the reason for reduced activity is due to destabilization of the active motif. Further studies into this drastic change in fluorescence enhancement could have implications in the design of structure-switching FE RNA considering **R3-10** is highly functional yet seems to be near a metastable equilibrium.

Sequences **R3-2**, **R3-8**, and **R3-9** exhibited higher affinity than **MGA**. This is especially impressive due to **MGA**(reported $K_d = 117$ nM [55], measured $K_d = 61.0 \pm 13.4$ nM) having an already extremely high affinity in comparison to other small-molecule aptamer which usually have dissociation constants in the μM range. Further optimization of these sequences to yield improved ϕ_F could be useful for *in vivo* or single-molecule imaging of RNA considering the high non-specific background and cytotoxicity of MG at high concentrations. Binding isotherms and dissociation constants of the highest affinity (**R3-8**), brightest (**R3-B-4**), and most abundant (in the Round 3 and Round 3-R pools, **R3-1**) sequences are plotted alongside **MGA** in Figure 2.6.

From the characterized sequences, only **R3-B-4** was measured to be brighter than **MGA** - exhibiting an 81% increased fluorescence enhancement. This is notable because **R3-B-4** was measured to have a K_d of 579 ± 76nM which is

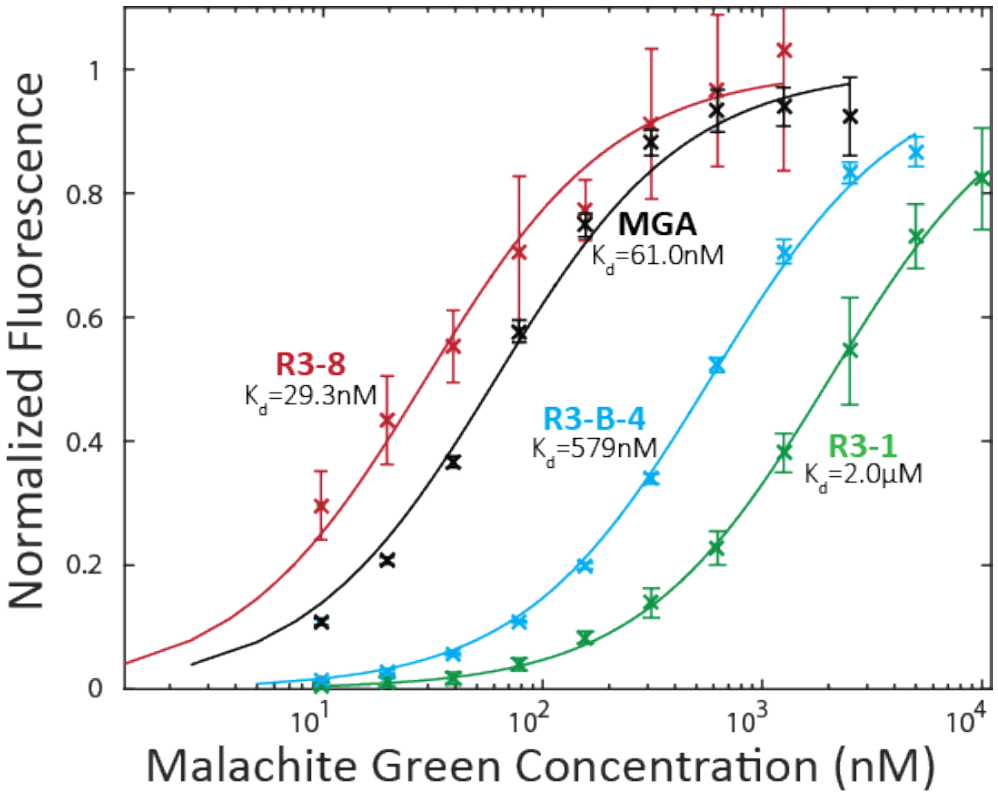

Figure 2.6: Measured dissociation constants for select sequences alongside **MGA**. **R3-8** represents the lowest K_d measured, Sequence **R3-B-4** represents the brightest aptamer discovered, and Sequence **R3-1** was the most abundant aptamer in the Round 3 and Round 3-R pools. Measurements were performed at 100 nM RNA.

nearly 10-fold higher (lower affinity) than **MGA**. We confirmed this quantum yield measurement with the technique previously described[55]. After performing an RNA titration at 500 nM malachite green, we obtained a ϕ_F for **R3-B-4** of 0.348 (Figure 2.7 B). This yields a fluorescence enhancement of 4390-fold, nearly identical to that obtained using the malachite green titration. Interestingly, using an RNA titration yielded a dissociation constant around 10-fold higher than measured using

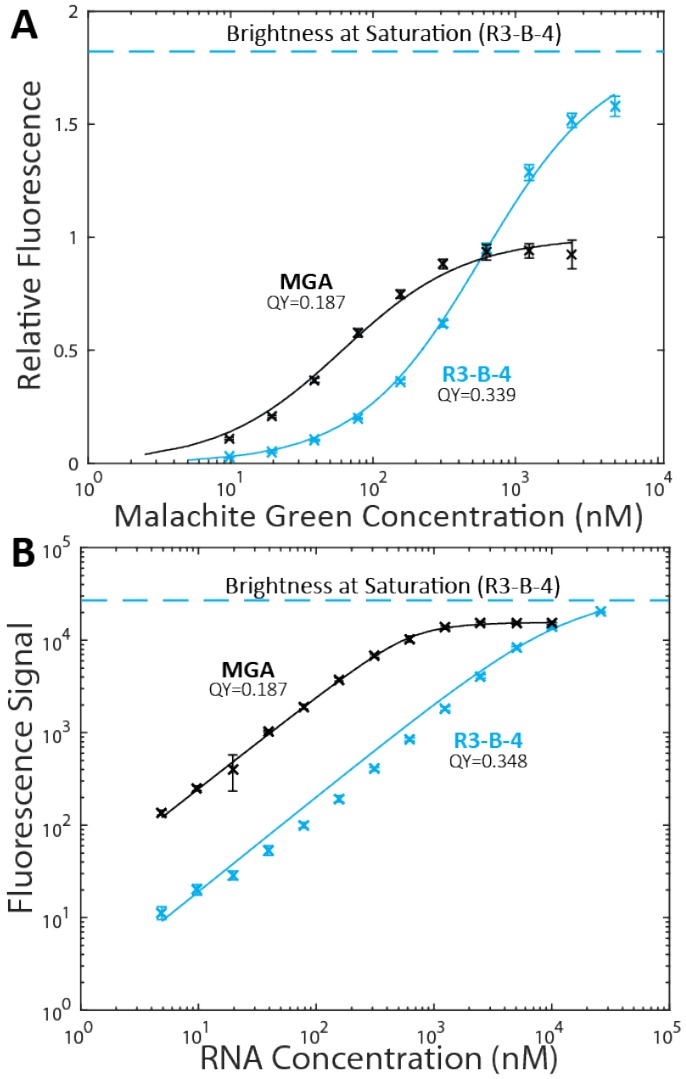

Figure 2.7: Sequence **R3-B-4** exhibits a higher quantum yield under two different assay conditions compared to the brightest known fluorescence-enhancing aptamer in literature **MGA**. **A)** Performing an MG titration with **R3-B-4** resulted in a quantum yield of 0.339 (FE = 4270-fold) based on relative fluorescence to **MGA**. Measurement was performed at 100 nM RNA. **B)** Performing an RNA titration confirmed the reported quantum yield of 0.187 (FE = 2360-fold). Sequence **R3-B-4** was measured to have a quantum yield of 0.348 (FE = 4390-fold). Measurement was performed at 500 nM MG.

a MG titration. We believe this results from stabilized dimerization of **R3-B-4** resulting in a reduced folding efficiency. Because ϕ_F is measured in this assay under conditions of limited dye, the disparity in measured K_d should have no effect on the accuracy of the determined quantum yield.

Across all characterized aptamers, there exists little correlation between binding affinity and fluorescence enhancement ability. Likely, this is because FE is dependent on certain binding modes which can each span a range of K_d values. This emphasizes that affinity-based SELEX methods for identifying FE aptamers can yield aptamers of sub-optimal functionality. For example, a sequence such as **R3-B-4** would likely be lost over many rounds of SELEX due to competition by dimmer but higher-affinity sequences (e.g. sequence **R3-8**). This is because, in SELEX, FE aptamers are enriched based on an unrelated property of the aptamer (low K_d). On the other hand, GRAP display can gear the selection towards identifying aptamers with both of these properties through reducing target concentration (selection for high affinity) or by increasing the fluorescence threshold that triggers sorting (selection for high FE).

This ability is well demonstrated by plotting the brightness and K_d of all sequences (Figure 2.8). We see that high affinity, low-brightness sequences (which would quickly come to dominate a pool using traditional SELEX) exhibit negative enrichment from Round 3 to Round 3-B and Round 3-R. For example, Sequence

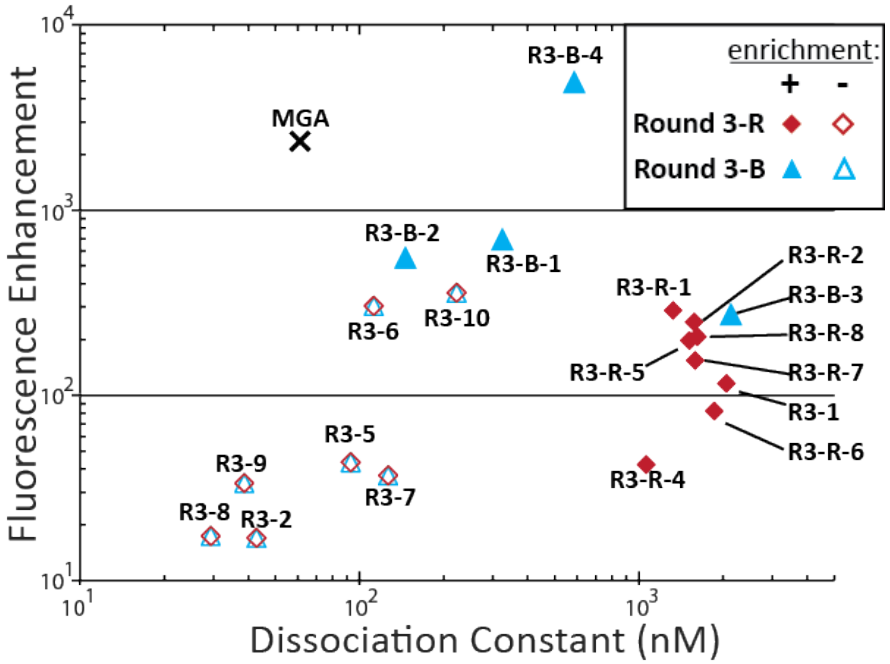

Figure 2.8: Plotting the dissociation constant and FE of each aptamer shows that GRAP display can effectively enrich sequences that would otherwise be lost with traditional SELEX. We see that many high affinity sequences are negatively enriched (unfilled shapes) going into the Round 3-R (red diamond) and Round 3-B (blue square) pools. In SELEX, these high affinity sequences would continue to dominate a pool regardless of their FE performance.

R3-5, which represents approximately 10% of the Round 3 pool, experiences a 55-fold reduction in copy number in Round 3-R and 13-fold reduction in copy number in Round 3-B, despite being of high affinity (K_d = 92.6 nM).

2.2.7 Direct Selection of Unique Fluorescence Profiles

It has been shown that different fluorophores can exhibit distinct emission profiles upon binding the same dye target[16,109]. We were able to use GRAP display

to actively select for aptamers based on the profiles they emitted. To do this, we collected all fluorescing particles above the primary excitation peak of MG (618nm) over the first two rounds of selection. Then, in Round 3, we plotted the fluorescence signals generated across two distinct emission wavelengths (670±30nm and 780±60nm) which were then sorted into blue-shifted (Round 3-B) and red-shifted (Round 3-R) subpopulations (Figure 2.4).

Characterization of these pools confirmed that GRAP display allows us to isolate aptamers that generate multiple distinct emissions profiles upon binding to the same dye. While we initially expected to identify only two distinct profiles between the Round 3-B and Round 3-R pools, we instead found that the aptamers could be grouped into a total of 11 unique spectral families based upon their shared excitation and emission profiles (Table 2.3).

As expected from the varied FACS plots (Figures 2.3 and 2.4), we measured the excitation and emission profiles for all sequences in Round 3 and found the emission maxima spanned a wide range of wavelengths (from 649nm 675nm). Upon separating Round 3, however, we see that the sequences positively enriched in Round 3-B have a much narrower range of emission maxima (649 654nm). Likewise, all positively-enriched sequences in Round 3-R share a unique red-shifted emission maximum at 675nm. The excitation and emission spectra for the most

Name	λ_{exc} (nm)	λ_{em} (nm)	**Spectral Family**
R3-1	646	675	1
R3-2	641	669	2
R3-3	634	652	3
R3-4	631	649	4
R3-5	635	658	5
R3-6	633	651	6
R3-7	635	658	5
R3-8	628	651	7
R3-9	641	669	2
R3-10	634	652	3
R3-B-1	631	650	8
R3-B-2	634	652	9
R3-B-3	631	649	4
R3-B-4	636	654	10
R3-R-1	646	675	1
R3-R-2	646	675	1
R3-R-3	636	656	11
R3-R-4	646	675	1
R3-R-5	646	675	1
R3-R-6	646	675	1
R3-R-7	646	675	1
R3-R-8	646	675	1
WG	634	653	

Table 2.3: Excitation and emission spectra were measured for the 22 characterized sequences and the maximum wavelengths for each recorded. Spectral Family 1 exhibited the highest red-shifted emission, as expected from this family's high representation and enrichment in the Round 3-R pool. Fluorescence from sequence R3-3 was observable with the spectrophotometer used and is thus recorded.

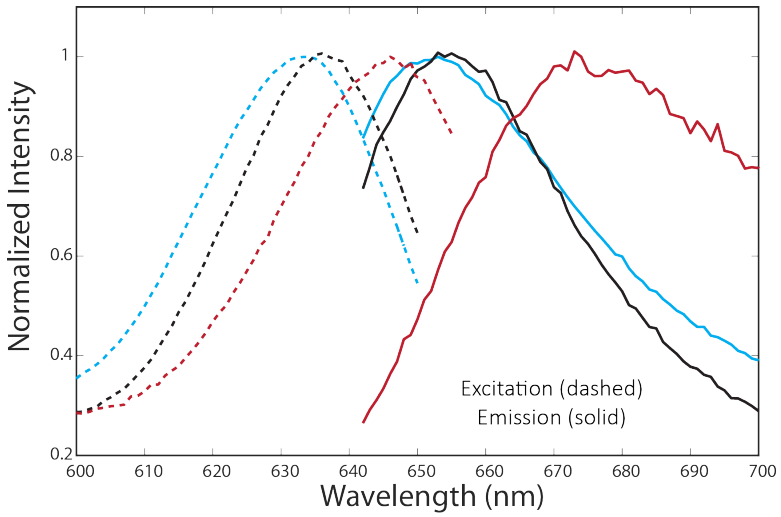

Figure 2.9: Excitation (dashed) and emission (solid) spectra of selected sequences. **R3-4** (blue) is the most abundant sequence in the Round 3-B pool, while **R3-1** (red) is the most abundant sequence in the Round 3-R pool. These sequences also represent the furthest blue- and red-shifted emission profiles of any discovered. For comparison, **MGA** (black) is plotted.

abundant sequences in the Round 3-B and Round 3-R pools are shown in Figure 4 alongside **MGA**.

Upon inspection with mfold software, we predicted that the 675nm emission spectra shared by members of Spectral Family 1 was because of a conserved motif among these sequences. We performed minimization experiments around a predicted hairpin and discovered a short, 15-nucleotide (nt) motif was responsible for the unique red-shifted fluorescence spectrum(**tR3-R**, Figure 2.10). To our knowledge, this is the shortest FE motif described in literature. Short motifs are advantageous for *in vivo* tagging of RNAs that might not tolerate a larger RNA tag. Based on the presence of this motif, we expect that 98 out of the top

100 most enriched sequences in the Round 3-R pool share this same spectrum, exemplifying the abilities of GRAP display to differentiate aptamers based upon their fluorescence emission spectrum.

2.2.8 *In Vivo* Expression of Aptamers in Mammalian Cells

Malachite green is regarded to be highly cytotoxic[168–170] and has been shown to fluoresce brightly and non-specifically *in vivo* due to intercalation into the cell wall[109]. However, an esterized variant of malachite green (MG-e, Figure 2.11 A) is known to be membrane-permeable and has been successfully used for *in vivo* imaging[171]. We performed a screen of our aptamers against MG-e and found that, while most sequences exhibited minimal or no FE of MG-e, those containing the **tR3-R** motif exhibited significantly increased fluorescence with MG-e compared to the same concentration of MG(Figure 2.11 B). Due to its improved FE with MG-e, we wished to express the **tR3-R** motif in cells to demonstrate that aptamers selected using GRAP display can successfully be applied in vivo.

Previously, Filonov et al. had found that the tRNA aptamer scaffold[172] is rapidly targeted for degradation by endonucleases[173] in mammalian cells. For this reason, they adapted a previously known RNA three-way junction scaffold to more stably express RNA aptamers *in vivo*. With the re-engineered F30 scaffold, they

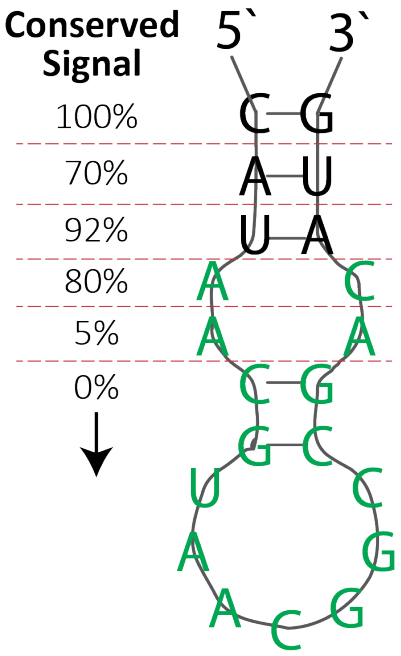

Figure 2.10: Minimum motif responsible for emission at 675nm. A truncation experiment was performed at 200 nM RNA and 1μM MG for each motif by pair-wise deletion in the hairpin. Shown in green is the motif (**tR3-R**) necessary for FE of MG. To our knowledge, this represents the shortest fluorescence enhancing motif described in literature.

showed stable expression of Broccoli-tagged RNA in HEK293T cells. Additionally, incorporating more than one Broccoli motif into the scaffold yielded a substantial increase in fluorescence compared to a single Broccoli motif. We wished to adapt this scaffold for our use and so we incorporated 3 copies of **tR3-R** into the F30 scaffold (Figure 2.11 C). Next, we transfected the resulting F30-3x-tR3-R motif into HEK293T cells using the pAV U6+27 plasmid (See Methods).

We confirmed expression of our transcript using flow cytometry by observing a $\approx$ 40% increase in the mean fluorescence of the transfected (containing F30-3x-tR3-

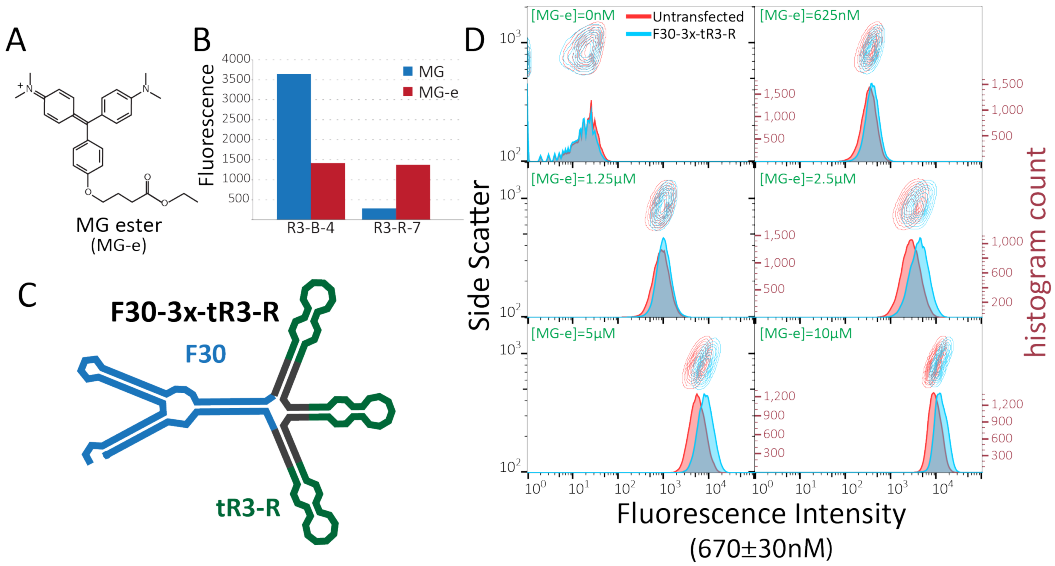

Figure 2.11: A) Chemical structure of malachite green ester (MG-e). MG ester (MG-e) is known to be membrane-permeable and suitable for *in vivo* fluorescence imaging. **B)** While many sequences such as **R3-B-4** exhibited significantly decreased activity with MG-e compared to MG, sequences such as **R3-R-7** (which contains the **tR3-R** motif) from the Round 3-R pool exhibited significantly increased activity. Measurements were performed at 400 nM RNA and 1M MG or MG-e. **C)** The predicted secondary structure of the F30-3x-tR3-R transcript showing the three tandem tR3-R repeats. **D)** F30-3x-tR3-R enhances the fluorescence of cells. Flow cytometry analysis shows ≈40% increased fluorescence between transfected (blue) and untransfected (red) cells at concentrations above 1.25M. 30,000 events were recorded at all concentrations.

R) versus the untransfected cell lines at concentrations above 1.25 μM (Figure 2.11 D). This data suggests that, given a suitable scaffold, aptamers selected *in vitro* against a suitable dye target could be readily applied *in vivo*. Future selections could benefit by performing the selection directly against a suitable cell-permeable dye that exhibits low non-specific fluorescence *in vivo*.

2.3 Conclusion

Fluorescent proteins are immensely important in cellular biology and their discovery remains an active field of research. However, there exist very few methods for detecting an equally important biomolecule: RNA. One common method involves tagging an RNA of interest with a FE aptamer that activates fluorescence of a non-fluorescent chromophore. However, very few FE aptamers exist that can operate in this capacity. This is because the method of discovering FE aptamers relies on enriching for sequences based on their binding affinity to the target chromophore. which is a poor method of discovering new fluorescence enhancing aptamers because the FE abilities of many discovered aptamers is found to correlate poorly with binding affinity[108,109]. Thus, a new manner of aptamer discovery that allows selection based on both binding affinity and fluorescence enhancement is needed.

Gene-linked RNA Aptamer Particle (GRAP) display answers this need while also providing a number of advantages over other selection methodologies. By selecting for fluorescence abilities rather than strictly binding affinity, GRAP display allowed us to discover the brightest fluorescence-enhancing aptamer known even though the sequence exhibits significantly lower binding affinity than other aptamers in the pool. Perhaps most interesting was the ability to isolate a

population of red-shifted fluorescence-enhancing aptamers and identify a short, 15nt RNA motif responsible for the shared emission profile. The ability to "tune" the fluorescence of resulting aptamers would be of great use in developing a catalog of aptamers yielding unique fluorescence profiles and based upon a small number of related chromophore dyes.

We show that by directly selecting for fluorescence rather than binding affinity, we can isolate RNA aptamers that are considerably more functional than **MGA** after just three rounds of screening. Importantly, we demonstrate that we can measure the fluorescence signal of every GRAP across multiple emission windows, enabling us to directly identify aptamers on the basis of their spectral emission. We believe the versatility of this method and applicability towards any suitable fluorescent target should greatly expand the toolbox of reagents available for studying functional RNA.

2.4 Methods

2.4.1 Immobilization of malachite green

We immobilized malachite green (MG) isothiocyanate (Molecular Probes) on the surface of 2.7 μm Dynabeads M-270 amine beads (Invitrogen). The beads were activated in 100 mM NaHCO3 buffer, pH 8.5, and suspended in 100 mM NaHCO3

buffer, pH 9. MG isothiocyanate was added to the beads at a final concentration of 500 μg/mL and incubated with rotation overnight. The supernatant was removed and the beads were washed thoroughly using dimethyl sulfoxide (DMSO, Fisher Scientific) with gentle heating to remove free MG. Washing was repeated $\approx$10 times until no visible color change of the DMSO was observed. After washing, the beads were visibly darkened, indicating successful functionalization though no quantification was possible. Washed beads were suspended in TNaTE buffer (10 mM Tris pH 7.5, 150 mM NaCl, 0.01% Tween-20, 0.1 mM EDTA) and stored at 4°C.

2.4.2 DNA preparation

The single-stranded DNA library and all primers used were synthesized by Integrated DNA Technologies. Modified oligos were purified using HPLC, while all others underwent standard desalting. Randomized bases were hand-mixed at a 25:25:25:25 ratio. The DNA library (Library) consisted of molecules containing a randomized 40nt region flanked by conserved PCR primer-binding sequences. The reverse primer (RP) was purchased alongside a 25-nt poly-T tail (RPpT) at the 5′ end for use with PCR and emulsion PCR, respectively. The forward primer (FP) contained a T7 promoter region to allow transcription of the expressed aptamers.

To functionalize carboxylic acid beads with primer, we purchased FP and the 25nt poly-T capture strand featuring a 5′-amino modification separated from the first nucleotide by a 72-atom internal spacer (AmM-FP, AmM-pT). We purchased a FAM-labeled FP complementary sequence (FAM-FPc) and FAM-labeled poly-A (FAM-pA) to determine the extent of primer expression on the bead surface. We also purchased a FAM-labeled poly-T RP (FAM-RPpT) for use with 1 cycle of PCR to fluorescently-label the displayed DNA after emulsion PCR cleanup. We purchased an Alexa Fluor 647-labeled FP (647-FP) to determine extent of RNA coating upon emulsion transcription. The methods by which these processes were performed, and the sequences for ordered oligos are given below.

Oligo	Sequence ($5' \rightarrow 3'$)
Library	TAATACGACT CACTATAGGG ACACAATGGA CGNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNTAACGGCC GACATGAGAG
FP	TAATACGACT CACTATAGGG ACACAATGGA CG
RP	CTCTCATGTC GGCCGTTA
RPpT	TTTTTTTTTT TTTTTTTTTT TTTTTCTCTC ATGTCGGCCG TTA
FAM-RPpT	/56-FAM/-TTTTTTTTTT TTTTTTTTTT TTTTTCTCTC ATGTCGGCCG TTA
AmM-FP	/5AmMC6//iSp18//iSp18//iSp18//iSp18/-TAATACGACT CACTATAGGG ACACAATGGA CG
AmM-pT	/5AmMC6//iSp18//iSp18//iSp18//iSp18/-TTTTTTTTTT TTTTTTTTTT TTTTT
647-FP	/5Alex647N/-TAATACGACT CACTATAGGG ACACAATGGA CG
FAM-FPc	/56-FAM/-CGTCCATTGT GTCCCTATAG TGAGTCGTAT TA
FAM-pA	/56-FAM/-AAAAAAAAAA AAAAAAAAAA AAAAA

2.4.3 Coupling forward primer to particles

500 μL of 1 μm MyOne carboxylic acid beads (Life Technologies,10^{10}/mL) were activated by washing once with 500 μL of 0.1 M NaOH and five times with 500 μL ddH2O. The beads were incubated with 1X conjugation buffer (200 mM NaCl, 100 nM imidazole in water), 100 μM AmM-FP, 100 μM AmM-pT, and 250 mM 1-ethyl-3-(3-diemthylaminopropyl)carbodiimide (EDC) in 50% v/v DMSO overnight at RT with rotation. The beads were then washed twice with 500 μL 0.1

M 2-(n-morpholino)ethanesulfonic acid (MES) buffer (100 mM, pH 4.7) (Pierce Biotechnology) for 10 minutes with rotation and resuspended in MES buffer.

To passivate unreacted functional sites, we converted unreacted carboxyls into amino-reactive NHS-ester groups in the presence of 250 mM EDC and 100 mM N-hydroxysuccinimide (NHS) suspended in MES buffer for 30 minutes. The beads were conjugated with 20 mM amino-PEG12 in MES buffer for one hour. The FP-modified beads were then washed two times with TT buffer (0.05% Tween-20, 50 mM Tris, pH 7.5) for 5 minutes to quench the reaction, resuspended in 500 μL TNaTE buffer, and stored at 4 °C.

To test conjugation efficiency, we incubated two separate batches of 0.2 μL of FP beads with 1 μM FAM-FPc and 1 μM FAM-pA, respectively, in 100 μL TNaTE buffer at RT for 5 minutes. Each batch of beads was washed once with 100 μL TNaTE buffer, and analyzed using a BD FACSVerse cytometer (BD Biosciences). To confirm that the FP beads were suitable for PCR, we performed 20 cycles of PCR with 1 μM FP, 1 μM FAM-RPpT, 1 nM template DNA, 1X GoTaq PCR Master Mix (Promega), and 1 μL FP beads in 20 μL total volume. These beads were then washed three times using 100 μL 100 mM NaOH, and washed once with 100 μL TNaTE buffer. These beads were resuspended in 100 μL TNaTe buffer and analyzed on the FACSVerse. Generally, the fluorescence signal after PCR

was ≈10-fold lower than for the primer conjugation efficiency test described above which we suspect is a result of steric interference on the surface of these beads.

2.4.4 RNA library pre-enrichment

We PCR amplified $1x10^{14}$ sequences of the starting library with 1 μM FP, 1 μM RP, and 1X Master Mix to prepare a double-stranded library for use in transcription. To minimize PCR bias, we performed only 6 cycles of PCR on a total volume of 8 mL. We performed PCR with the following conditions: 98 °C for 2 minutes, followed by 6 cycles of 98 °C for 15 seconds, 57 °C for 30 seconds, and 72 °C for 30 seconds. The DNA was separated from the reaction mixture using phenol:chloroform extraction and the aqueous phase was concentrated via butanol extraction. The recovered DNA was ethanol precipitated using 2.75 volumes of ethanol and 0.1 volume 3 M sodium acetate (pH 5.5), and resuspended in TE (10 mM Tris, 0.1 mM EDTA, pH 7.5) buffer.

Transcription of $8x10^{14}$ of these DNA sequences was performed in five reaction volumes with the TranscriptAid T7 High Yield Transcription Kit (Thermo Scientific), following the manufacturers protocol. After DNase treatment for 1 hour at 37 °C, the recovered RNA was purified using a 6% urea-TBE acrylamide gel (National Diagnostics), cut from the gel by UV shadowing, and recovered

through the crush-and-soak method in water. The eluent was ethanol precipitated as described above, and the recovered pellet was suspended in TE buffer.

In the first round of SELEX, we heat denatured ≈1x10^{15} RNA molecules for 5 minutes and quickly placed on ice to cool. The RNA was incubated with ≈ 10^9 MG-coated beads in PBSMT buffer (1X PBS pH 7.4, 5 mM $MgCl_2$, 0.01% Tween-20), washed two times with PBSMT, eluted using PBSMT buffer heated to 95 °C, and reverse-transcribed using cloned AMV reverse transcriptase (Invitrogen) following the manufacturers instructions. We amplified the recovered DNA using PCR and transcribed into RNA as described above for use in a second round of pre-enrichment. In the second round of SELEX, we reduced the amount of RNA and target, using $4.5x10^{14}$ RNA sequences and $6.5x10^7$ MG-coated beads.

2.4.5 GRAP synthesis

We generated our GRAPs using emulsion PCR followed by emulsion transcription. The oil phase in both reactions consisted of 95% mineral oil, 4.5% Span 80, 0.45% Tween 80, and 0.05% Triton X-100. For emulsion PCR, the aqueous phase consisted of 1X PCR Master Mix (Promega), 10 nM FP, 1 μM RP-polyT, ≈ 10^8 beads, and 1.5 pM template DNA. Water-in-oil emulsions were generated for emulsion PCR by adding 1 mL of the aqueous phase to 7 mL of oil phase in a DT-20 tube (IKA), which was locked into an Ultra-Turrax Device (IKA). The addition was

performed drop-wise over 30 seconds while being stirred at 620 rpm for 5 minutes. We performed emulsion PCR under the following cycling conditions: 95 °C for 5 minutes, followed by 40 cycles of 95 °C for 30 seconds, 57 °C for 60 seconds, and 72 °C for 60 seconds.

After PCR, the emulsions were collected into an emulsion collection tray (Life Technologies) through brief centrifugation. The emulsions were broken using 3 mL of 2-butanol and transferred to a 50 mL tube. The particles were pelleted by centrifugation at 4,000 x g for 5 minutes and the supernatant was decanted. The particles were resuspended in emulsion breaking buffer (BB) (100 mM NaCl, 1% Triton X-100, 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA) and transferred to a 1.5 mL tube. The beads were sonicated, vortexed, centrifuged for 15 seconds at 13,000 x g and placed in a magnetic separator (MPC-S, Life Technologies). The supernatant was removed via pipette and the beads were washed once with 100 μL PBS, once with 100 μL BB, and three times with 100 μL 100 mM NaOH to generate ssDNA and remove non-covalently bound DNA. We confirmed that 20$\pm$10% of the particles contained DNA by annealing $\approx$50,000 beads with 1 μM FAM-RP-poly-T at RT for 5 minutes. The particles were washed once with 200 μL PBS-T and analyzed using the FACSVerse (BD Biosystems).

To generate double-stranded DNA on the particles, we subjected them to one cycle of PCR using 1X GoTaq PCR Master Mix and 1 μM FAM-labeled poly-T

RP. The beads were washed twice with 100 μL BB and resuspended in 500 μL TNaTE.

Emulsion transcription was performed using the T7 High Yield Transcription Kit. The aqueous phase consisted of 0.4X TranscriptAid Reaction Buffer, 4 mM of each rNTP, 5 mM $MgCl_2$, $\approx 5x10^7$ DNA-coated beads/mL, and 0.4X TranscriptAid Enzyme mix. For the starting (Round 0) pool, we used 1 mL of aqueous phase and 7 mL of oil phase, and the emulsions were prepared in the UltraTurrax device as described for the emulsion PCR process. In every subsequent emulsion transcription reaction (Rounds 1-3), the emulsions were made in quadruplicate through drop-wise addition of 50 μL aqueous phase over 15 seconds to 450 μL oil phase in 2 mL cryogenic vials (Corning), which were being mixed at 1000 rpm by Teflon flea micro stir bars (Fisher). The reaction vials were stirred for 3 minutes, then incubated at 37 °C for 6-10 hours.

The emulsions were kept on ice and broken using 3 mL of cold 2-butanol and transferred to a centrifuge tube. The particles were pelleted by centrifugation at 4°C and the supernatant was decanted. The particles were then resuspended in cold BB and transferred to a 1.5 mL tube. The particles were sonicated, vortexed, centrifuged for 15 seconds at 13,000 x g, and then placed on an MPC-S. The supernatant was removed by pipette. Further wash steps were performed without centrifugation using cold BB, until all oil in the supernatant was removed. To

confirm that the particles contained RNA, we incubated ≈50,000 GRAPs with 1μM 647-FP at RT for 5 minutes, washed once using 100 μL TNaTE, and analyzed with the FACSVerse cytometer.

2.4.6 FACS screening

Prior to screening, we incubated FP beads with malachite green (Spectrum Chemical) at concentrations ranging from 10 nM to 10 μM in PBSMT pH 7.4 to determine levels of background and non-specific binding. This is an appropriate baseline, because ≈80% of particles after ePCR contain no PCR product and thus little to no RNA, making them indistinguishable from FP beads. We performed FACS on the Aria II instrument (BD Biosystems), isolating particles that exhibited a fluorescence signal above background (defined as the sort gate) due to MG activation and which also exhibited a strong fluorescence signal from the FAM-labeled dsDNA. This ensured that only particles which expressed both RNA and its parent DNA were collected and enriched.

We incubated $\approx 10^8$ particles in round 1 and $\approx 10^7$ particles in all later rounds. In each round, we performed a similar MG titration as described for FP beads above, and chose an MG concentration such that $\approx 0.1\%$ of the total population fluoresced in the reference gate. All incubations were performed at a bead concentration such that $\approx 5,000$ events/second were registered on the FACS instrument. Because of

our interest in aptamers that cause MG to fluoresce at different wavelengths upon excitation, we set reference gates for the following bandpass filters: 670±30nm, 730±45nm, and 780±60nm. We collected any events that fluoresced higher than the background signal, which was established as described above, and the sorted particles were directly added to a PCR mixture and amplified for use in later rounds. Over the course of two additional rounds of selection, the MG concentration was decreased 100-fold. After the second round of selection, we observed no further enhancement in binding ability of the Round 3 pool at 10 nM MG. We then separated the R3 pool into R3-B and R3-R pools by incubating the beads at a concentration of 1 μM MG, separating using the FACS instruments, and PCR amplified as described.

Channel:	**670±30nm**	**730±45nm**	**780±60nm**
Round 1:	6,646	6,713	4,946
Round 2:	7,532	5,536	5,210
Round 3:	21,332	18,795	18,640
Round 3-B:	≈20,000	0	0
Round 3-R:	0	0	≈20,000

2.4.7 Sequencing analysis

High-throughput sequencing was performed at the Stanford Functional Genomics Facility (SFGF) using the MiSeq sequencer (Illumina). Using a custom MATLAB

script implementing the Bioinformatics Toolbox, the sequence data was analyzed to obtain copy numbers of all sequences. For all sequences that were present in >20 copies in the raw sequencing data, we calculated the reads-per-million (RPM). The RPM was then used to calculate a sequence "enrichment" value. Enrichment is defined as the fractional increase in copy number relative to the previous round. Thus, a sequence that went from 100 RPM in the Round 2 pool to 500 RPM in the Round 3 pool would have an enrichment value of 5 in the Round 3 pool. Enrichment values were only calculated for sequences with >200 copies in the current round.

We chose to characterize the union of the top 5 sequences by RPM, and the top 5 sequences by enrichment for each pool. In total, 22 unique sequences were identified in this manner: 10 in the Round 3 pool, 4 in the Round 3-B pool, and 8 in the Round 3-R pool. The DNA sequences containing T7 promoter regions on the $5'$ end were synthesized at Stanfords Protein and Nucleic Acid (PAN) facility. They were then PCR amplified, transcribed, and purified using the methods described above. The RPM and Enrichment values for all 22 sequences in the given rounds are given in Table 2.4. n.d. means not detected (<20 copies in the raw data).

Name	RPM Round 3	RPM Round 3-B	RPM Round 3-R	Fitness Round 3	Fitness Round 3-B	Fitness Round 3-R
R3-1	157406	8281	571285	6.34	0.05	3.63
R3-2	111565	11302	95829	13.01	0.10	0.86
R3-3	110502	13082	1417	6.04	0.12	0.01
R3-4	96745	707584	2473	8.43	7.31	0.03
R3-5	95075	7517	1740	10.04	0.08	0.02
R3-6	2295	574	130	25.90	0.25	0.06
R3-7	165	n.d.	n.d.	22.08		
R3-8	131	n.d.	n.d.	21.29		
R3-9	195	n.d.	175	20.92		0.90
R3-10	3249	2173	142	17.19	0.67	0.04
R3-B-1	26	1102	n.d.		42.67	
R3-B-2	25	422	n.d.		17.20	
R3-B-3	106	1213	n.d.	8.44	11.45	
R3-B-4	172	1493	n.d.	4.30	8.67	
R3-R-1	5609	n.d.	28654	1.89		5.11
R3-R-2	7201	n.d.	18279	0.38		2.54
R3-R-3	14488	n.d.	8753	1.93		0.60
R3-R-4	47	n.d.	708			15.16
R3-R-5	18	n.d.	249			13.61
R3-R-6	7	n.d.	82			11.44
R3-R-7	39	n.d.	399			10.26
R3-R-8	8	n.d.	76			10.17

Table 2.4: Reads-per-million (RPM) and fitness (enrichment) values for the sequences obtained after high-throughput sequencing. n.d. means that the sequence was not detectable above 20 copies in the raw data. No entry in fitness means that the fitness value was uncomputable due to the sequence being undetectable in either the current or previous round.

2.4.8 Measurement of aptamer fluorescent properties

The following sequence was ordered, amplified using a suitable primer set, transcribed, and purified to yield **MGA** which was used for direct comparison of sequence performance: 5′-TAATACGACT CACTATAGGG AACACTATCC GACTGGCACC CCCCTGCCAG GTAACGAATG AAGTGCTTTT CTCGATCTCG TGACCCGCGC ACTAGTCGCG AAGGTGTATG TCCTTGGTCA TTAGGATCC-3′.

A stock MG concentration was prepared in PBS-T buffer (1X PBS pH 6.1, 0.01% Tween-20) and subjected to serial dilutions from 20 μM to 20 nM in the same buffer. The concentration of MG was measured using a NanoDrop spectrophotometer (Thermo Scientific) and calculated with an extinction coefficient of 148,900 cm-1/M at 618 nm[162] after allowing MG to equilibrate overnight in PBS-T. Assays were prepared in flat-bottomed 96-well MicroFluor 2 plates (Thermo Fisher Scientific) at a final concentration of 100 nM RNA, 5 mM $MgCl_2$, and 0.01% Tween-20, with MG concentrations ranging from 10 nM to 10 μM. Wells were excited using the Tecan M1000 Plate Reader at 630±20 nm, and emission was measured at 675±20 nm. For each MG concentration measured, an RNA-free standard was measured and subtracted from the signal as a baseline. The signals were plotted at each MG concentration and least-squares fit to a 1:1 complexation model using a custom MATLAB script according to Equation 2.1 as described.

S_λ values were obtained by comparing the integral of the excitation and emission spectra for each aptamer to that of **MGA** across the regions used for excitation (610−650 nm) and emission (655−695 nm) measurements. Excitation/emission profiles were obtained as described below and fit to a polynomial using MathCad (Parametric Technology Corporation).

We confirmed the obtained ϕ_F for **R3-B-4** by performing an RNA titration in the PBS-T buffer at 5 mM $MgCl_2$. We prepared each RNA concentration at a final concentration of 500 nM

MG and performed serial dilutions of **MGA** from 10 μM to 10 nM. For **R3-B-4** we measured fluorescence at an RNA concentration of 26 μM and performed a serial dilution from 10 μM to 10 nM. This was done using the Tecan M1000 and measured under the described conditions. The data was fit to the equation previously described [55].

2.4.9 Measurements of excitation and emission profiles

Solutions containing 1X PBS, 5 mM $MgCl_2$, 3 μM MG, and 1 μM RNA at pH 6.1 were prepared for each aptamer, and excitation and emission spectra were measured on a UV spectrophotometer in a quartz cuvette (Hellma Analytics). The measurements were performed in triplicate, and a base-line corrected average was obtained after subtracting an RNA-free standard from the measurement.

2.4.10 Minimization of red-shifted motif

A total of 7 sequences were characterized that exhibited a unique, red-shifted emission profile with a maximum at 675nm. These sequences (Spectral Family 1) were found to have a conserved motif near the 3′ end of the random region and likely included parts of the reverse primer region based on mfold software [166,167]. The predicted hairpin was minimized by removal of single base pairs and by measuring the fluorescence response. Each motif variant was prepared at a concentration of 500 nM and incubated with 1 μM malachite green in PBSMT buffer

at pH 6.1. The minimization experiment showed that a short, 15nt motif (5′-AACGUAACGGCCGAC-3′) is minimally required for activation of malachite green while still exhibiting this unique red-shifted emission profile. This hairpin is significantly shorter than the active motif identified from **MGA** and could find use in the development of aptamer switches, *in vivo* expression, or in RNA origami where smaller sequences are preferred.

2.4.11 Molecular cloning of aptamer expression vectors

To generate vectors for expression of **R3-B-4** and the scaffolded **tR3-R** motif (Figure 2.11 B) under the control of a U6 promoter, the pAV-U6+27 plasmid[174] was purchased from Addgene (catalog 25709). The DNA sequences for **R3-B-4** and **tR3-R** were synthesized as gBlocks (Integrated DNA Technologies) with a 5′ SalI digestion site and a 3′ XbaI digestion. The cloning strategy and point of insertion was conducted as previously described[173].

2.4.12 Mammalian cell culture

HEK 293 cells were thawed into a 10cm tissue culture plate coated with 0.1% gelatin in DMEM/F12 buffer (Thermo Fisher) supplemented with 10% Fetal Bovine Serum. Once ≈90% confluent, the cells were dissociated using 0.25% Trypsin EDTA (Thermo Fisher) and $5x10^5$ cells were reverse transfected with 50 g of

plasmid using Lipofectamine 2000 (thermo Fisher) in 6 well plates according to the manufacturers specifications. 48 hours post-transfection, the cells were dissociated using 0.25% Trypsin EDTA and resuspended in Tyrodes solution (136 mM NaCl, 5 mM KCl, 2 mM CaCl2, 1 mM MgCl2, 10 mM glucose, 10 mM HEPES, pH 7.4) containing MG-ester before being analyzed with FACS.

2.4.13 Effect of internal spacer on primers

It should be noted that we found that a 72-atom spacer is necessary to separate the amino coupling from the first nucleotide in both the FP and poly-T primer sequences that are displayed on the FP beads. When these oligonucleotides were conjugated to the beads without a spacer, PCR efficiency was dramatically reduced; we believe this results from steric effects. This was further tested by conjugating an amino-dPEG24-t-butyl ester (Quanta BioDesign) to the surface of the beads using EDC, deprotecting the spacer by incubating three times with 100% trifluoroacetic acid (TFA) for 30 minutes, and then using EDC to conjugate amino-modified primers containing no internal 72-atom spacer to the surface of the beads before lastly passivating any unreacted carboxyl groups in the manner described above.

We compared the PCR efficiency of 1) beads functionalized with a 72-atom internal spacer, 2) beads displaying dPEG24 on the surface coupled to amino-modified primers, and 3) beads displaying no 72-atom spacer nor dPEG24 and

coupled to amino-modified primers. We found that the PCR efficiency of 1) and 2) were similar, while we were unable to display DNA on the surface of 3) (data not shown).

Next, we determine the PCR efficiency of 4) beads displaying both i) amino-modified poly-T primer with no internal spacer, and ii) a dPEG24 coupled to amino-modified FP with no internal spacer; and 5) beads displaying both i) amino-modified FP with no internal spacer, and ii) a dPEG24 coupled to amino-modified poly-T with no internal spacer. We found that we were able to display DNA on the surface of 4) but not 5) (data not shown). This indicates that the FP was occluded in 5) likely due to steric effects. For this reason, we always used a 72-atom spacer between the surface of the beads and all of the displayed primers.

2.4.14 Effect of poly-A tail on aptamer performance

During GRAP display, newly-transcribed RNA sequences incorporate a $3'$ 25-nt poly-adenosine (poly-A) tail that enables hybridization to capture strands on the surface of the FP beads. Though the poly-A tail and poly-T capture strand are exactly the same length, the poly-A tail does not need to be fully complementary with the poly-T capture strand in order for the RNA to be immobilized. Thus, it is possible that regions of the poly-A tail may therefore play a role in aptamer folding. For this reason, dissociation constants and brightness values were determined for

all 23 aptamers both with and without the poly-A tail regions. Having assessed the performance of both the poly-A(+) and poly-A(-) versions of each sequence, we have reported data for only the more functional of the two variants of each aptamer. It is important to note that future selections using GRAP display must be aware that the poly-A tail may effect the performance of the final aptamer, though the effects of the poly-A tail were often very minimal.

We had also attempted to immobilize the RNA using a capture strand of defined sequence (i.e. not a poly-A) that was complementary to a region of the reverse primer in the RNA strand. Likely due to steric effects and the inability for the RNA tail to "burrow" into the displayed primers, we were unable to coat FP beads that were prepared as described above with RNA. For this reason, we used a poly-A/poly-T capture mechanism for all of our described experiments.

Chapter 3

Re-evolving Known Aptamer Motifs for Improved Function

3.1 Introduction

In nature, species adapt to environments that apply a selective pressure which favors the survivability of some individuals over others. Usually, these changes are so slight as to have almost no effect on survival of an individual. However, on rare occasions, an inheritable difference between individuals can result in an advantageous trait that, over multiple generations, can come to dominate a gene pool. This is the principle behind both natural and directed evolution.

Most DNA sequences that are present in a random pool used for SELEX are indistinguishable from one another under selective pressure - they neither perform nor survive. However, some of these contain a functional gene, or "motif," that allow them to survive in the given environment and, over many rounds of selective pressure, come to dominate a pool. Oftentimes, these functional motifs are identified upon SELEX convergence because of sequence similarity between otherwise very different aptamers.

In SELEX, the designed libraries generally contain from $20 - 80$ random bases, yielding a theoretical sequence diversity of $10^{12} \approx 10^{48}$. Under normal laboratory conditions, a population of fewer than 10^{15} sequences ($\approx$ 1 nanomole) are able to be tested. Thus, only a minute subset of total sequence space is generally explorable using directed evolution. While short functional motifs may occur multiple times in a population of this size, highly functional sequences generally comprise multiple interacting motifs and are thus exceedingly rare[64]. For example, the $preQ_1$ riboswitch is the shortest aptamer known in nature at 34nt[97]. The sequence complexity of a motif this long is $4^{34} = 3x10^{20}$. It is these more longer, more complex, and increasingly rare motifs which are desired from a selection as we search for more functional aptamers, as increased complexity means more opportunity for an aptamer to interact with and act upon its target in unique and interesting ways.

Thus, the biggest concern with any particle-based selection techniques is that approaching a sequence diversity rivaling that of SELEX is not possible. This is because the throughput of FACS is only $\approx 10^8$ per hour. In these approaches, other methods of reducing library diversity are required, such as limiting the total amount of diversity by reducing the number of random bases, or by performing a handful of SELEX rounds prior to initializing the display method. In PD, for example, 2 rounds of traditional SELEX are generally performed in order to reduce the diversity of the starting library to a throughput reconcilable with FACS. However, for many targets of interest, functional pre-existing aptamers already exist, and performing an entirely new selection may not yield meaningful improvement, or can fail spectacularly due to the biases previously discussed (Section 1.6.2).

In this Chapter, we apply a method of reducing the starting diversity of a random library needed to find functional aptamers. To do this, we destabilize and then segmentally randomize a known functional motif and perform three rounds of GRAP display to yield brighter and higher-affinity sequences than the best described malachite green aptamer. This method works by biasing a library towards functionality, and using directed evolution to identify novel sequences that perform well. Even with a theoretical diversity far in excess of what is explorable with GRAP display, we identify sequences from our library that perform far better than those discovered in Chapter 2. Importantly, this is done without any pre-

enrichment using traditional SELEX showing that GRAP display can be easily applied to optimize known aptamer sequences. Next, we discuss the rationale for our current design, and propose design parameters for future mutagenized libraries that can optimally explore the sequence space around a known motif using any method of directed evolution. Lastly, we discuss the applications of this method for yielding improved aptamer sequences for fluorescence-enhancing and other functional aptamer motifs.

3.2 Results and Discussion

3.2.1 Re-Evolving a Known Motif for Improved Function Using GRAP Display

MGA-m contains two short helices which are divided by a central, asymmetric, 11 base bulge (Figure 3.1). Both helix 1 (containing the 5′ and 3′ ends of the sequence) and helix 2 are not considered to be responsible for binding MG, rather their primary role is stabilizing the binding pocket[175]. For this reason, we decided to destabilize helix 1 in the design of our library. Particularly, we were interested in evolving additional bases in this region that, through tertiary interaction or structural stabilization, could yield better performance. This could result from stabilized

(lengthened) helix stem region of **MGA-m** or through tertiary interactions of the scaffold with the functional pocket of **MGA-m**.

First, we removed 8 bases from **MGA-m** that are necessary to stabilize the MG-binding pocket. Next, we inserted the reduced motif into the center of two 15-base random regions (Figure 3.1). We wished to show that the functional performance could be improved through additional stability or tertiary interactions. In this library, the exact **MGA-m** sequence should be fully present in about 1:65,536 (1 sequence in 4^8)

We next performed three rounds of GRAP display as previously described by first subjecting this library to two consecutive emulsion reactions that transform it into a library of GRAPs. These were then incubated with 50 nM malachite green and analyzed using FACS. Because we expect stabilization of the conserved **MGA** motif to yield fluorescence-enhancing aptamers with identical excitation and emission profiles, only particles that fluoresced above background in the 670 ± 30 emission window were collected. The brightest particles were then PCR amplified for the next round of selection. This process was repeated and the concentration of free MG was reduced from 50 nM at the start, to 1 nM in Round 1, and 500pM in Round 2. Round 3 GRAPs were prepared as described, but no further enrichment was observed at concentrations below 500pM MG and we deemed this pool converged (See Methods).

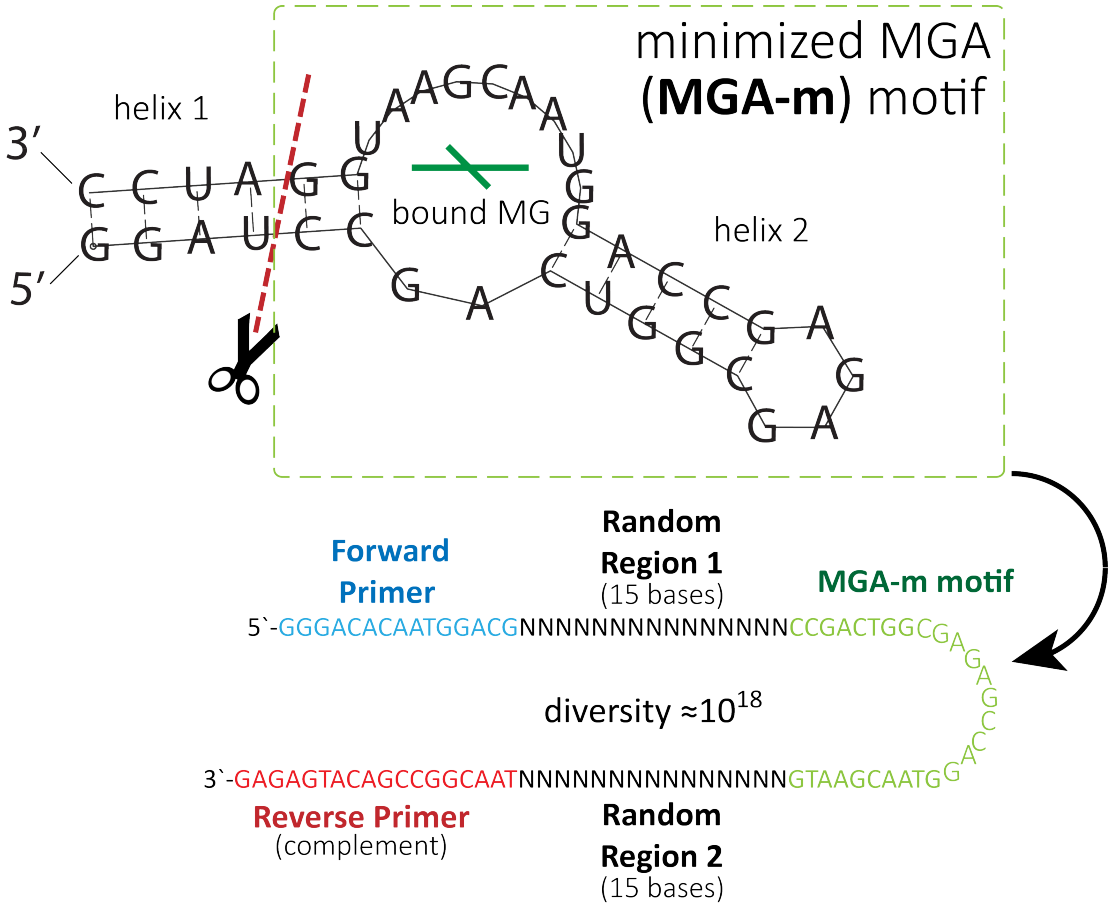

Figure 3.1: A) Secondary structure of **MGA-m** showing active binding pocket. Boxed is the region that was inserted into a starting library and flanked by a total of 30 random bases.

Name	**Sequence ($5' \rightarrow 3'$) of core region** conserved MGA-m region is bolded
MG-D1	ACACGCCATTGTCCT**CCGACTGGCGAGAGCCAGGTAACGAATG**AGGTTTCTAAGTGTA
MG-D2	ACACGCCATTGTCCT**CCGACTGGTGAGAGCCAGGTAACGAATG**AGGTTTCTAAGTGTA
MG-D3	CAGATAATCGGATTC**CCGACTGGCGAGAGCCAGGTAACGAATG**GGCCGGTTTCTGTGT
MG-D4	CCTTTAAGGGGCTCG**CCGACTGGCGAGAGCCAGGTAACGAATG**CGAGGCTCGGAAAAG
MG-D5	CGCCCAGGATCTCTG**CCGACTGGCGAGAGCCAGGTAACGAATG**CAAGACGTCCCCTGT
MG-D6	CGCCTCCAATTTGG **CCGACTGGCGA AGCCAGGTAACGAATG**CCAATAGCCCCTCCG
MG-D7	CGTACCATAAGGCCT**CCGACTGGCGAGAGCCAGGTAACGAATG**AGGGGATTAACACTC
MG-D8	CGTACCATAAGGCCT**CCGACTGGCGAGCGCCAGGTAACGAATG**AGGGGATTAACACTC
MG-D9	CGTACCATAAGGCCT**CCGACTGGCGAGGGCCAGGTAACGAATG**AGGGGATTAACACTC
MG-D10	GGTCTTAATGGTAAG**CCGACTGGCGAGAGCCAGGTAACGAATG**GCACGCGACAAACCA

Table 3.1: 10 sequences were identified from HTS for further characterization. Sequences were not found to be effected by a poly-A tail, and were characterized without. In bold is the conserved region of the library. Due to random PCR error, some mutations and deletions occur in this region.

At this point, we performed HTS of the Round 2, and Round 3 pools and $2-4\times10^6$ sequences were identified from each pool. Using a custom MATLAB script and the Bioinformatics Toolbox, we filtered and ranked all sequences based on their copy number (abundance) and fractional increase in copy number over the previous round(enrichment or "fitness") as described in Section 2.2.4. We were particularly interested in sequences with an enrichment value greater than 1 (which were thus represented more than in the previous round) because previous studies have shown that enrichment can be a better predictor of sequence performance than its abundance[132,134]. After this analysis, we chose the five most abundant sequences and the five most enriched sequences in Round 3 for a total of 10 sequences that were subjected to further characterization (Table 3.1). We characterized **MGA-m** and **MGA** alongside these 10 sequences in order to see how well a segmentally randomized library can "re-evolve" the function of **MGA-m**.

3.2.2 Fluorescence Performance of Identified Sequences

Sequences were characterized using the methods previously described (Section 2.2.6 and Methods Section 2.4.8). In calculating K_d and ϕ_F, S_λ was set to 1 for all sequences because different excitation and emission profiles were not observed. This is expected given the conserved binding mode (same binding pocket) of MG with these sequences and **MGA**.

Of the 10 sequences identified through HTS and sequence analysis, every one exhibited fluorescence enhancement of malachite green. **MG-D10** and **MGA-m** (the 38-base minimized motif) both exhibited affinities above the concentration range probed in our characterization so dissociation constants were unable to be reported. The remaining sequences all exhibited similar or higher affinity than **MGA**. Two sequences, **MG-D3** and **MG-D8**, showed improved FE than **MGA** (Figure 3.2). This is noteworthy because previous *in vivo* attempts at mutating a FE aptamer did not result in improved FE[110]. Though the K_d value calculated for **MGA** differs from that previously measured (2.2.6), they are within error of one another (measured $K_d = 103.0 \pm 31.0$ compared to previously measured $K_d = 61.0 \pm 13.4$). K_d and FE values are shown in Table 3.2 and plotted in Figure 3.2.

MG-D7 was the most abundant sequence in Round 3 indicating that it was most consistently enriched over the 3 rounds of GRAP display. Interestingly, the sequence exhibiting the highest enrichment, **MG-D8**, differs from **MG-D7** by only a single nucleotide and in the conserved region of **MGA-m**. This mutation was evidently introduced as a random PCR error but provided a significant enough performance enhancement so as to be highly enriched during the sorting process. Furthermore, **MG-D8** was found to exhibit a $\approx$ 4-fold quantum yield increase over **MG-D7** at the cost of a $\approx$4-fold decrease in affinity. This is interesting

Name	K_d (nM)	**Fluorescence Enhancement** (fold increase)	ϕ_F
MG-D1	63.6 ± 6.2	1638 ± 37	0.130
MG-D2	59.6 ± 5.4	1445 ± 30	0.115
MG-D3	44.4 ± 19.8	2799 ± 270	0.222
MG-D4	60.2 ± 17.1	1338 ± 87	0.106
MG-D5	42.9 ± 9.8	1262 ± 62	0.100
MG-D6	55.6 ± 16.5	1224 ± 81	0.097
MG-D7	42.4 ± 7.1	1064 ± 38	0.084
MG-D8	112.4 ± 53.9	4602 ± 565	0.365
MG-D9	30.3 ± 4.5	995 ± 30	0.079
MG-D10	n.d	n.d	n.d
MGA	103.0 ± 31.0	2360 ± 179	0.187

Table 3.2: Characterization results of identified sequences. Equations 2.1 and 2.3 were used to calculate the K_d, fluorescence enhancement (fold increase over background), and ϕ_F of each aptamer. Reported are the mean and standard deviation of samples prepared and measured in triplicate.

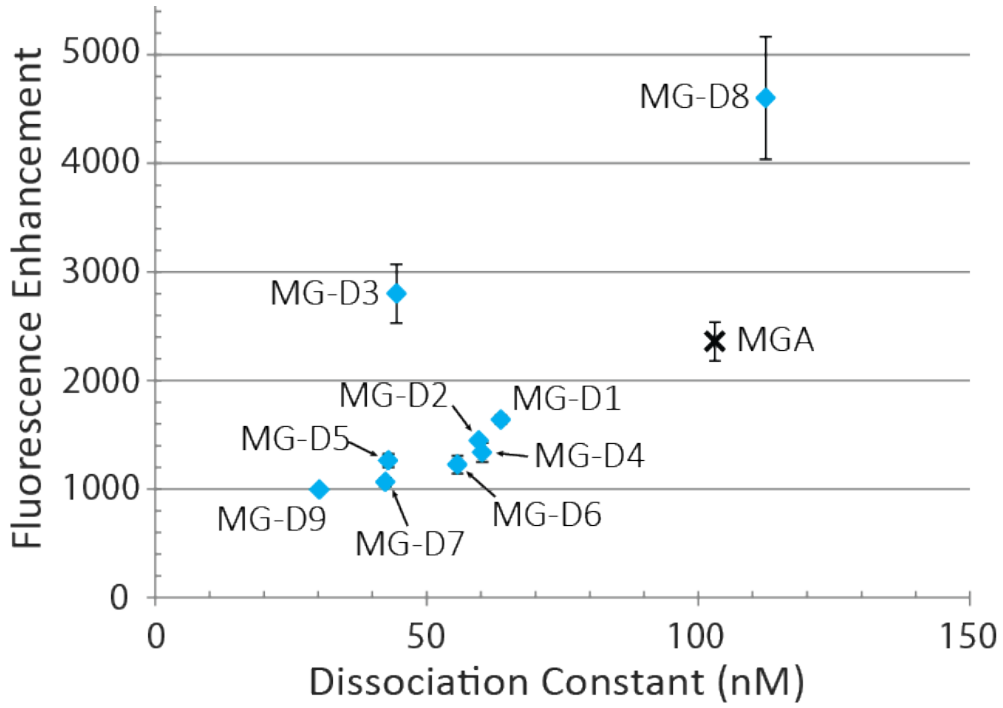

Figure 3.2: K_d and FE (relative to background) of aptamers discovered using GRAP display through segmentally randomizing **MGA-m**. Sequences **MG-D8** and **MG-D3** were measured to be brighter than **MGA**, with **MG-D3** exhibiting a significantly improved affinity. Error bars are representative of standard deviation from measurements prepared and measured in triplicate.

because it shows a single nucleotide is responsible for both increased quantum yield and a decreased affinity in the same aptamer. This grants further evidence that fluorescence enhancement has little or no correlation to affinity and that only GRAP display can effectively and reproducibly isolate functional aptamers even under a vast abundance of higher affinity sequences.

3.2.3 Stability improves Performance

While accurate prediction of tertiary structure is beyond the capabilities of available technologies, we can predict the secondary structure of the discovered sequences and suggest reasons for improved performance. For example, mfold structures of many characterized aptamers show what look to be significantly stabilized stem regions of the functional motif. Sequence **MG-D3** (Figure 3.3) has a helix 1 region with 17 predicted base pairs (up from 4 in **MGA-m**), which is stable and long enough to adopt a more complex tertiary structure. This could explain both its greatly enhanced fluorescence and high affinity compared to **MGA** and **MGA-m**. Previous studies using RNA Spinach have shown that correcting mismatches and bulges in the functional region can lead to improved affinity and brightness[141]. Possibly, **MG-D3** could be made to perform better through correcting or deleting the mismatched pairs of the full structure.

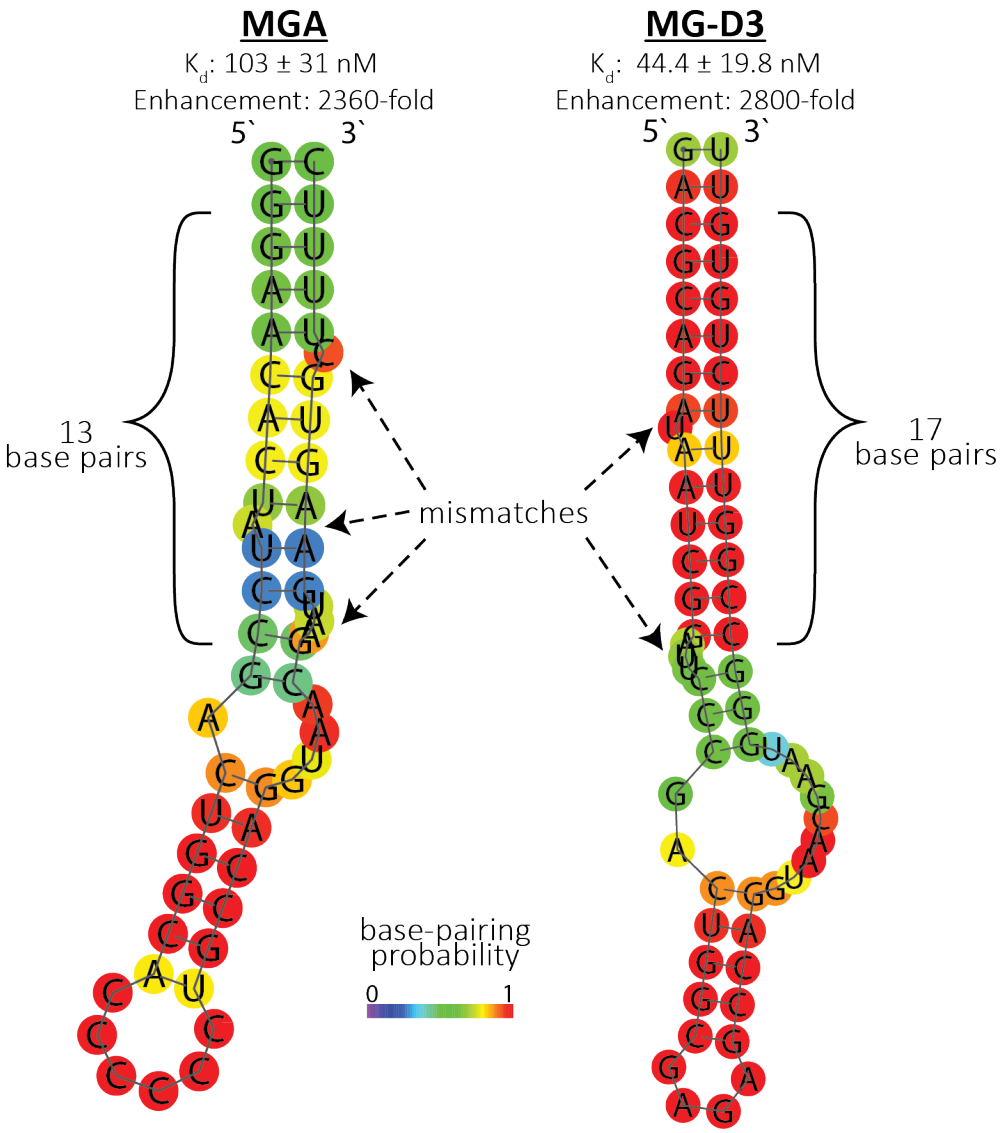

Figure 3.3: Predicted mfold folding of the active regions of **MGA** and **MG-D3** including nearby stabilizing regions. In this structure prediction, **MGA** was truncated by 58 nts on the 3′ end, while **MG-D3** is truncated by 11 nts on the 5′ end and 17 nts on the 3′ end. **MG-D3** is expected to have a longer and more stable stem region than **MGA** which could result in the improved fluorescence enhancement and binding affinity towards MG. **MGA**[107] contains a different hairpin motif than **MGA-m**[55] which was used in the design of our library.

3.2.4 Mutations in Conserved Motif Results in Modified Function

Aptamers **MG-D7**,**MG-D8**, and **MG-D9** are distinct from one another by a single base mismatch (X) at the same position in the four-base loop of helix 2 (Figure 3.4. In **MGA-m**, helix 2 was identified as a GNRA-type tetraloop through crystallization[175]. GNRA-type tetraloops are named for the sequences which form them (i.e. 5′-GNRA-3′). Here, N can be any base (A, U, C, or G), while R is a purine base (either A or G).

GNRA-type tetraloops fold to yield an unusual G:A base pair between their first and fourth bases. Additionally, in all described tetraloops, the "N" nucleobase is stacked with the "R" nucleobase. In the crystal structure of **MGA-m**, however, the A16′ of GAG<u>A</u> is flipped out of place, compared to the canonical GNRA-type tetraloop position[175]. Instead, it is predicted to interact with the preceding G15′. In their analysis, Baugh et al. predicts that the flipping of A16′ is an artifact of crystallization. They also predict that the helices and GNRA-tetraloop appear to play no role in affecting the aptamer active site. The results of this selection seem to contradict these claims and could be worth pursuing in later studies.

The most stable predicted structure for **MG-D7**, **MG-D8**, and **MG-D9** all suggest a misfolded binding pocket (Figure 3.4). Nonetheless, **MG-D7** contains

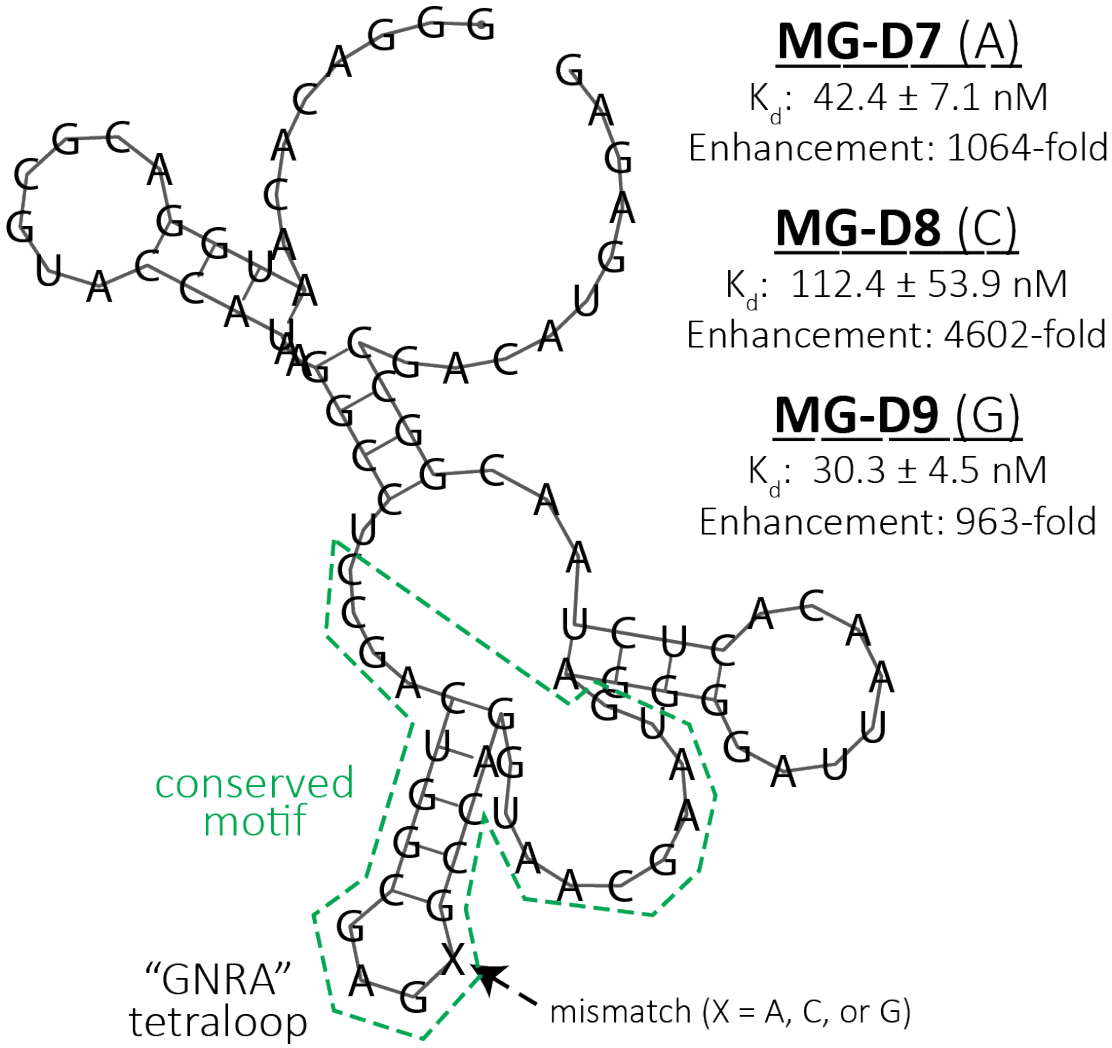

Figure 3.4: Predicted folding of **MG-D7**, **MG-D8**, and **MG-D9** which differ from one another by a single base mismatch (X) yet have wildly varying activity. This suggests that the identified base plays an important role in either 1) stabilizing the binding pocket of **MGA-m**, or 2) interacting with MG in the binding pocket.

containing GAG<u>C</u>-tetraloop. This change results in a 4-fold decrease in binding affinity, and 4-fold increase in MG enhancement.

While it is possible that the GAGC-tetraloop in **MG-D8** merely has a destabilizing effect on the structure and results in an unknown interaction that increases fluorescence enhancement of MG, this does not explain the insensitivity to a GAGG-tetraloop. The crystal structure of this motif suggests that A16′ undergoes base-base stacking with the preceding base, yet this was dismissed as a crystallization artifact. However, G-G stacking is a very common phenomenon in G-quadruplex, and favorable G-G stacking would explain the similar performance of **MG-D9** and **MG-D7**. If A16′ was not undergoing base-base stacking as the crystal structure implies, we would also expect a change in activity with a GAGG-tetraloop due to disruption of the G:A base pair between the first and last bases of the tetraloop. To confirm these suspicions, an interesting experiment would be to compare the performances of all N16' point mutations for **MGA-m** alone.

3.2.5 Statistics of Library Mutagenesis

As seen above and elsewhere[63,110,147,152], point mutations inside a motif can yield new and improved function. In the search for highly functional aptamers discovered using particle display, future selections could benefit from doping (introducing random point mutations into) a library to explore all single-, double-, triple-, etc. mutants of a known fluorescence enhancing motif. In the ideal case, this mutagenized library would allow us to fully sample the maximum edit distance from

the known motif. Here, we discuss a mathematical model on the distribution and optimization of a mutagenized library[151], how it has been previously applied[110], and a manner by which to improve it for future library design and particle-based selection performance.

The diversity of a starting library is dependent on the number of random bases and the distribution of these bases. Usually, libraries are synthesized with an equal chance that a random base (N) is A, T/U, C, or G. Then, the maximum diversity D of most libraries is dependent on length n in the following manner:

$$D(n) = 4^n \tag{3.1}$$

If we want to explore sequence space around a known motif, we need to begin with a motif of length n, and apply a mutation rate m to every base. With a library of correct design, we could then explore all single-, double-, triple-, etc. ($p = 1, 2, 3...$) point mutations of this motif. In this example, the distribution of sequences with p point mutations is given by the binomial coefficient:

$$\binom{n}{p} = \frac{n!}{p! \cdot (n-p)!} \tag{3.2}$$

Starting with a known motif, each base can be independently mutated into one of 3 other bases. Specifically, the total number of sequence variants with p mutations from the starting motif is given by

$$N_p(n,p) = (3)^p \binom{n}{p} \tag{3.3}$$

We wish to design a library such that the maximum number of point mutations is fully explorable. Thus, we need to determine the distribution of sequences in a library as a function of a given mutation rate m. Here, we assume at a given m there is an equal likelihood for mutating into any of the other 3 bases. That is, at a mutation rate of $m = 15\%$, there is a 5% chance of a base mutating into each of the 3 other bases, and an 85% chance of the same base remaining unchanged. The distribution of a mutagenized library as a function of motif length n, mutation rate m, and number of mutations p follows the binomial distribution as given:

$$F_m(m,n,p) = (1-m)^{n-p} \cdot m^p \cdot \binom{n}{p} \tag{3.4}$$

For a random library containing N_{tot} sequences, the distribution of sequences is given:

$$N_m(m,n,p) = N_{tot} \cdot (1-m)^{n-p} \cdot m^p \binom{n}{p} \tag{3.5}$$

As has been derived previously[151], a mutation rate m and maximum number of fully-sampled point mutations p_{max} can be obtained by solving the system of equations:

$$(3)^p \binom{n}{p} \leq N_{tot} \cdot (1-m)^{n-p} \cdot m^p \binom{n}{p} \tag{3.6}$$

$$\frac{\partial}{\partial p} N_{tot} \cdot (1-m)^{n-p} \cdot m^p \binom{n}{p} = 0 \tag{3.7}$$

for m and p. To solve explicitly, the binomial function can be replaced by the gamma function ($\Gamma(n+1) = n!$) and the equations solved.

3.2.6 Optimization of Library Mutagenesis

Using the described manner of library mutagenesis, a selection was performed against DFHBI using the known RNA Spinach aptamer[110]. Filonov et al. designed a library by solving Equations 3.6 and 3.7 such that all mutants with up to $p = 8$ mutations were present (e.g. $n = 54$, $m = 14\%$, $N_{tot} = 10^{14}$ and $p_{max} = 8$. Note, the authors inexplicably used a reduced mutation rate from what is given by solving the system of equations). Next, five rounds of SELEX were performed to remove non-functional sequences. They then expressed this library into *E. coli*

cells and performed an *in vivo* sort on $\approx 10^3$ colonies using FACS to identify aptamers which exhibit fluorescence enhancement.

Aside from the enrichment limitations of SELEX in comparison to particle display (discussed in Section 1.6.2 and Section 1.8.2), this selection is also hindered by significantly over-representing low p sequences.

Reducing low p sequence redundancy is done by optimizing the design of the library such that 1) we can fully sample the highest number of point mutations (p_{max}); and 2) we have the maximum amount of sequence space covered for $p > p_{max}$. In other words, we want to sample the largest edit distance without unnecessarily sacrificing the overall diversity of our library by over-representing highly-conserved (low p) sequences. We can do this by removing the inequality in Equation 3.6 and solving for m, using p_{max} obtained from Equations 3.6 and 3.7 :

$$(3)^{p_{max}} \binom{n}{p_{max}} = N_{tot} \cdot (1-m)^{n-p_{max}} \cdot m^{p_{max}} \binom{n}{p_{max}} \tag{3.8}$$

In the example mentioned above, solving Equation 3.8 results in the recommended mutation rate increasing from 14% to 21%. Surprisingly, this still fully samples all sequences with $p \leq 8$ point mutations. Instead, the redundancy of unique sequences with $p = 8$ mutations is reduced from an average of 2.1 copies to only 1. This change also increases the number of point mutations in the average sequence from $n \cdot m = 7.5$ to 11 which allows a wider exploration of nearby sequence

space. Lastly, this reduces the percentage of the pool that is redundant from 42.7% to 7.5%. Because maximum enrichment for two sequences in SELEX is equal to the ratio of their K_d values[164], it is hugely important to reduce representation of sequences near $p = 0$ as these sequences are expected to maintain near full function and will out-compete rarer (on a per-molecule basis), highly mutated, better-performing aptamers which the selection is intended to find.

Modifying the mutation rate to include more of sequence space using traditional SELEX may yield only minimal practical improvements due to inherent limitations on enrichment[164]. However, it would greatly improve particle-based selections by increasing the breadth of sequence space explored. For example, in optimizing the same 54-base motif above using GRAP display ($N_{tot} = 10^8$) solving Equations 3.6 and 3.7 tells us that an initial optimal mutation rate of 8% allows us to explore all 1-, 2-, and 3-base point mutations. By solving Equation 3.8, we find that increasing the mutation rate to 22% still allows us to sample all 1, 2, and 3-base point mutations but with over-representation at $p = 0$ and $p = 1$ reduced significantly. Overall, this method reduces the total number of redundant sequences in the starting GRAP display library from 38% to only 0.3% (a 37.7% reduction in redundancy) (Figure 3.5).

While this may seem to bias our library towards highly mutated libraries where most sequences may not be functional, we are still able to fully sample all sequences

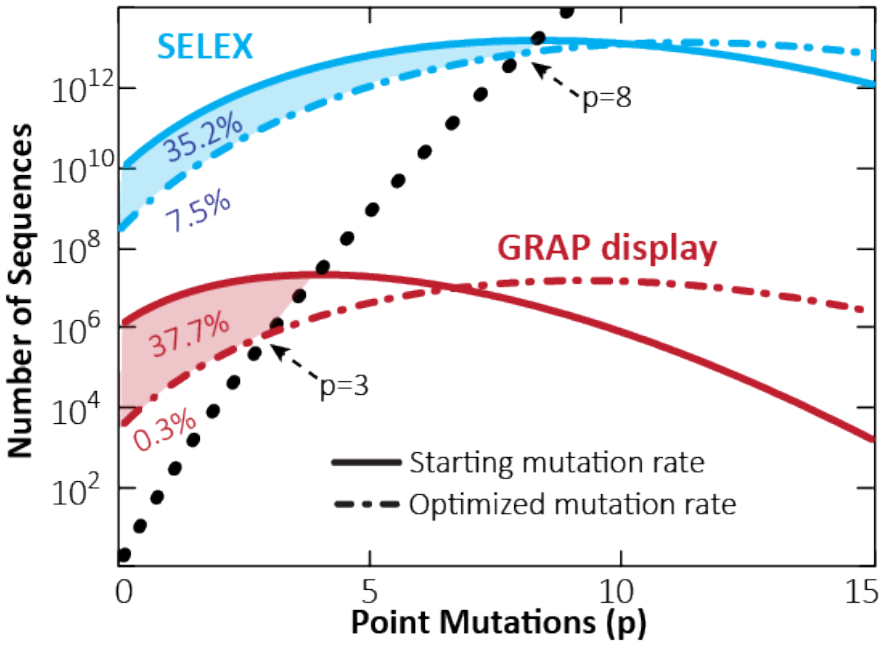

Figure 3.5: Originally suggested (solid) vs optimized (dashed) library design using SELEX (blue, $N_{tot} = 10^{14}$) and GRAP display (red, $N_{tot} = 10^{8}$) for the discussed motif (n=54nt). The total number of sequences possible with p mutations is given by Equation 3.3 (black dotted line). Shaded regions represent unnecessary sequence redundancy in the starting library. In the discussed example, increasing the mutation rate from 14% to 21% can reduce SELEX redundancy by 35.2% to only 7.5% of the starting library. For GRAP Display, increasing mutation rate from 8% to 22% reduces redundancy by 37.7% to only 0.3% of the starting library while still sampling all 1-, 2-, and 3-base point mutations.

with $p \leq 3$ mutations. Additionally, we expect that these high p sequences are more able to grant improved function to the starting motif due to their broader exploration of nearby sequence space.

Due to the phenomenal enrichment possible with particle-based display methods, we believe this manner of library mutation would still result in significantly improved sequences when compared with SELEX or *in vivo* screening. For example,

Filonov et al. discovered only 6 mutants from the starting SELEX library ($m = 14\%$) that exhibited high FE. 5 of these had $p \leq 3$ mutations from the 54-nt Spinach motif, while the remainder had $p = 4$ mutations. Using GRAP display, we would still be expected to sample 5 of these 6 clones from our starting library. In our view, limitations on SELEX enrichment only allows discovery of sequences with few point mutations due to their enormous starting (per sequence) abundance and high likelihood to function (and thus be collected during the partitioning process).

3.3 Conclusions

There exists a plethora of known aptamer sequences against a variety of targets, but oftentimes these sequences fall short of optimal or desired function. These shortcomings have lead to the development of numerous selection methodologies, but it is not always preferable to evolve a new aptamer for every desired function or application. For this reason, there have been a number of attempts at improving known aptamer sequences through incorporating mutagenesis[146,147] into the selection process using methods such as error-prone PCR[148], library design[151], or random splicing/recombination[149,150]. Sometimes, rationally mutating bases and segments of a given motif can improve their performance[136–138,140,141]. These methods allow for a deeper, more focused exploration of sequence space in the search for the most

functional aptamers. However, they often suffer from drawbacks, including the intractability of certain methods of mutagenesis[176], the limited number of features able to be displayed on a chip, and limitations in our computational and design capabilities[143–145]. A promising method involves designing a library around a known aptamer motif and allowing evolution to identify the optimized motif.

We wished to show that GRAP display can overcome physical throughput limitations in the search for improved functionality. First, we introduced a known fluorescence-enhancing aptamer motif **MGA-m** into a segmentally-randomized library and evolved aptamers that exhibited improved function over the best known malachite green aptamer **MGA**. This method works by making all sequences more resemble a known aptamer motif while providing enough diversity to invoke improved function. The resulting sequences showed a $\approx$ 100-fold increase in affinity and $>$ 10-fold increase in function over **MGA-m** alone, and a $\approx 2-5$-fold improved function (both affinity and FE) over the full-length **MGA**. We attribute these performance increases to highly-stabilized structural motifs. Interestingly, one sequence **MG-D9** exhibits significantly improved fluorescence enhancement capability and reduced affinity by introduction of a single base point mutation in a loop structure believed to be unrelated to MG binding. This result suggests that the tetraloop (which was previously classified as a GNRA-type[175]) may belong to a not-yet-described class of tetraloop structures. Lastly, we discussed a

known method of library design and introduced an optimization step in the process for designing mutagenized, or "doped," libraries. This optimization can more deeply explore sequence space around a known motif while reducing redundancy for applicability with particle-based display methods where throughput is limited.

3.4 Methods

GRAP display was prepared and performed in the manner described previously (Section 2.4). Buffer recipes and reagents remain as previously described. Additionally, sequences were ordered, synthesized, purified, and characterized as described previously.

3.4.1 DNA preparation

The single-stranded DNA library and all primers used were synthesized by Integrated DNA Technologies. Modified oligos were purified using HPLC, while all others underwent standard desalting. Randomized bases were hand-mixed at a 25:25:25:25 ratio. The DNA library (Library) consisted of molecules containing a conserved sequence (28bp) flanked by randomized 15nt regions which were in turn flanked by conserved PCR primer-binding sequences.

Identical primers were used as described previously (Section 2.4). Oligos used in this selection both for use in PCR and to transcribe the control sequences are as follows:

Oligo	**Sequence** ($5' \rightarrow 3'$)
Library	TAATACGACT CACTATAGGG ACACAATGGA CGNNNNNNNN NNNNNNNCCG ACTGGCGAGA GCCAGGTAAC GAATGNNNNN NNNNNNNNNN TAACGGCCGA CATGAGAG
MGA	TAATACGACT CACTATAGGG AACACTATCC GACTGGCACC CCCCTGCCAG GTAACGAATG AAGTGCTTTT CTCGATCTCG TGACCCGCGC ACTAGTCGCG AAGGTGTATG TCCTTGGTCA TTAGGATCC
MGA-m	TAATACGACT CACTATAGGG ATCCGACTGG CGAGAGCCAG GTAACGAATG GATCC

3.4.2 FACS screening

Prior to screening, we incubated FP beads with malachite green (Spectrum Chemical) at concentrations ranging from 1 nM to 100 nM in PBSMT pH 7.4 to determine levels of background and non-specific binding. This is an appropriate baseline, because ≈80% of particles after ePCR contain no PCR product and thus little to no RNA, making them indistinguishable from FP beads. We performed FACS on the Aria II instrument (BD Biosystems), isolating particles that exhibited a fluorescence signal above background (defined as the sort gate) due to MG activation and which also exhibited a strong fluorescence signal from the FAM-labeled

dsDNA. This ensured that only particles which expressed both RNA and its parent DNA were collected and enriched.

We incubated $\approx 10^8$ particles in round 1 and $\approx 10^7$ particles in all later rounds. In each round, we performed a similar MG titration as described for FP beads above, and chose an MG concentration such that $\approx 0.1\%$ of the total population fluoresced in the reference gate. All incubations were performed at a bead concentration such that $\approx 5,000$ events/second were registered on the FACS instrument.

We were not interested in collecting events that emitted at different wavelengths because we expect only a single emission wavelength. In the 670±30nm channel, we collected any events that fluoresced higher than the background signal, which was established as described above, and the sorted particles were directly added to a PCR mixture and amplified for use in later rounds. Over the course of two additional rounds of selection, the MG concentration was decreased 100-fold from 50nM in Round 0, to 1nM in Round 1, and 500pM in Round 2. After the second round of selection, we observed no further enhancement in binding ability of the Round 3 pool at concentrations below 500pM. The events collected in each round are as follows:

Channel:	670 ± 30**nm**
Round 1	11,839
Round 2	1,344
Round 3	1,473

3.4.3 Issues with contaminating sequences

GRAP display is incredibly efficient at enriching for highly functional sequences regardless of their starting copy number. To avoid issues with contamination stemming from this, it is highly recommended to change the primer sequences on the library after every new selection performed against the same target, assuming the first selection yields high affinity sequences. For example, the library (library 2) used in this chapter was ordered simultaneously as the library used in Chapter 2 (library 1). Library 2 was only used for a single pilot experiment, and that was several months prior to initializing the selection in Chapter 2. Nonetheless, sequences from library 2 were able to totally out compete library 1 which was realized upon sequencing. This lead to the failure of our first attempt at finding new MG aptamers from a näive library. This issue was later combated by performing less-stringent (higher target concentration) sorts such that we were collecting GRAPs at event rates above what would be expected from contaminating sequences.

Chapter 4

In Vitro Directed Evolution of Riboswitches with GRAP Display

4.1 Introduction

Upon completion of the Human Genome Project in 2003, it rapidly became apparent that only a small fraction $(1-3\%)$ of our genes directly code for proteins. A large portion of the remaining $\approx 99\%$, once termed 'junk DNA,' has since been found to be translated into RNA for a number of different purposes. For example, the Encyclopedia of DNA Elements (ENCODE) Project, which was announced in

2012[70], found that ≈ 75% of the genome is transcribed into RNA[71]; albeit much of it at very low levels[177,178].

The term non-coding RNA (ncRNAs) is simply RNA that does not undergo direct translation; and it encapsulates miRNAs (microRNA), rRNAs (ribosomal), lncRNAs (long non-coding), and numerous other types. While ncRNAs are becoming relevant in our understanding of cellular biology, the extent, importance, and purpose of their functions remains largely unknown. However, these RNA appear to comprise an additional layer of cellular signaling, control, and communication[179] beyond what is performed by proteins. Of those ncRNAs that are well described, they exist in roles as diverse as controlling subcellular structural organization[77,78]; promoting gene silencing[79,80], gene expression[81,82], and stem cell differentiation[83]; and determining cellular localization of macromolecules[84] in addition to numerous other functions[81,85,86].

Riboswitches are one manifestation of ncRNA that are often found in the 5′UTR (untranslated region) of bacterial or viral mRNA. When a riboswitch-specific ligand - typically a small molecule[96–99] - enters the local environment of these structures, the finely-tuned thermodynamics of folding shift so as to promote a structural reorganization and the modulation of downstream genes.

Typically, riboswitches are comprised of three domains: an RNA aptamer, a transduction region, and an active region[95]. In the natural environment,

the aptamer interacts with a target ligand which invokes a structural change in the transduction region. This results in the active region performing one of many diverse functions[95–97], such as translational control[180], transcription termination[181], mRNA degradation[180], alternative splicing[182], catalysis[183], or protein recruitment[184].

While *in vivo* functions of riboswitches are immensely important, *in vitro* applications are an equally exciting platform for their use. In these application environments, riboswitches are more often referred to as structure-switching aptamers to reflect that they are not RNA-specific. For example, structure-switching aptamers have been discovered that allow real-time sensing of small molecules[88,89], metal ions[90,185], and proteins[91,92]. As a relevant example, a structure-switching target-specific DNA aptamer has been utilized by our lab to measure[93] and then control[94] the levels of an important cancer drug - doxorubicin - in live animals. If more highly-functional, structure-switching aptamers were available, these technologies could become widely applied for arbitrary detection and *in vivo* control of numerous targets. This could have huge implications in the deployment of personalized medicine.

Because of their diverse display of function, riboswitches also have promise for use in gene-specific and targeted therapies, pharmaceuticals, diagnostics, molecular sensing, and cellular computing. However, new discovery of riboswitches is hindered

by their complexity (generally 100-200+ nucleotides in length[186]). This means that rational design and computational prediction methodologies can not adequately anticipate their function.

For example, a modular approach to riboswitch engineering yielded a number of *in vivo*-functional riboswitches responsive towards the small molecules SAM, lysine, guanine, FMN, and others. However, their switches were only amenable to a well-understood (and a mechanistically simply) type of function (rho-independent transcription termination) and yielded a lower dynamic range than natural switches[187]. Despite successful implementation with this design technique, rationally-designed switches do not typically resemble natural riboswitches which contain long, well-folded transduction regions and perform highly-efficient switching[97,188].

An alternative approach to rational or computation design of riboswitches is directed evolution. While some approaches have been applied to discover or engineer new riboswitches, they are usually only capable of screening 10^4 to 10^6 candidates. This is because they typically rely on *in vivo* expression platforms that come with new difficulties and are not amenable to truly high-throughput screening[189–192]. These *in vivo* methods suffer from drawbacks including heterogeneous riboswitch expression which necessitates additional normalization[110,193], the difficulty of performing effective negative or concentration-dependent selections[95], and the

incompatibility of riboswitches between different domains of life[194,195]. Meanwhile, *in vitro* methods have not been widely used as it is difficult to tether activity (switching) to a handle that allows preferential enrichment of functional switches.

With this understanding, we present a method of evolving riboswitches *in vitro* and using previously known aptamers as the recognition and active regions. We link these regions with a minimally-designed randomized transduction region, and perform a functional selection to enrich aptamers which become active only in the presence of their target ligand - adenosine triphosphate (ATP). We compare two methods of *in vitro* discovery using GRAP display and compare the discovered sequences with the best rationally-designed switches of the same class. We report a number of sequences which exhibit the desired behavior, and show that the most functional evolved random region does not follow the typical design parameters of riboswitches. These results suggest GRAP display could find widespread use in discovering RNA exhibiting novel function.

4.2 Results and Discussion

4.2.1 Design of a Riboswitch Library for the ATP-dependent Activation of MG

Previously, a DNA:RNA hybrid aptamer switch has been designed that allows ATP-dependent fluorescence-enhancement of malachite green using the known **MGA-m** motif. With this method, the authors were able to obtain a $\approx$ 4-fold enhancement of malachite green in the presence of 1 mM ATP when compared to the absence of ATP[126]. Other designs have shown controllable "turn on" fluorescence ($\approx 2 - 25$-fold enhancement) of chromophores with the addition of SAM, SAH, cyclic di-GMP, cyclic di-AMP, and cyclic AMP-GMP through incorporating rational design or *in vivo* evolution to modify existing function[101–103,196,197]. However, these switches are far from the dynamic "on"/"off" switches desired in application. Here, we intended to show that GRAP display is well-suited for the selection of functional riboswitches by designing an RNA library for *in vitro* selection and showing that this design is capable of resulting in the same reporting functions.

Generally, the process for designing a riboswitch involves first identifying a suitable target and reporter sequence. Then, these sequences are connected through a transduction region which incorporates elements of design in order to promote 1)

minimal interaction between this region and the aptamer regions, and 2) a delicate energy balance between the "on" (correctly folded) and the "off" (misfolded) state of the switch. Usually, design elements promote a metastable transduction region that is i) short (1-5base pairs), ii) contains "weak" base pairs like A-U and G-U, or iii)has totally mismatched base pairs[101]. In almost all cases, the "off" state is designed to be unfolded (non-interacting with either aptamer regions). In this way, binding the ligand can promote aptamer stabilization and yield correct folding of the reporter region. Lastly, several riboswitch designs are synthesized and characterized to determine the best-performing sequence.

In our approach, we wished to allow evolution to design a new riboswitch. Specifically, we intended to provide only a minimally designed transduction region and allow GRAP display to identify an aptamer whose two states are finely balanced such that it performs the desired switching function. We wanted to use this library to show ATP-dependent "turn-on" of malachite green only in the presence of ATP and hoped to identify a transduction region that interacts with the aptamer regions in the absence of target. While it goes against the principles of rational design, this behavior more closely mimics natural riboswitch. In this way, we are not introducing artificial constraints into the structure such as those that are implemented with rational design. To do so, we began with **MGA-m**

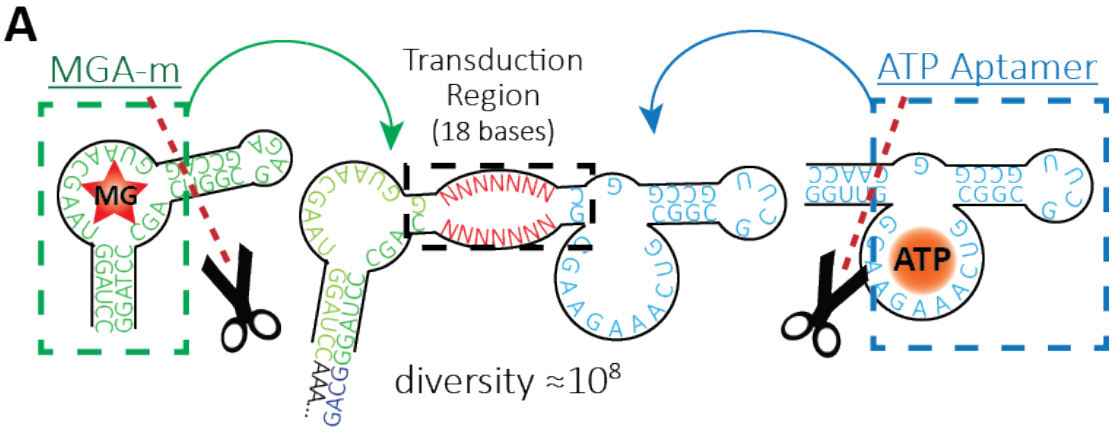

Figure 4.1: A) Our library was designed to incorporate the known **MGA-m** sequence and an ATP-specific aptamer. We removed stabilizing hairpins from both structures and connected them with a randomized transduction region.

and a known ATP aptamer[147], and replaced stabilizing structural regions of both motifs with a randomized transduction region (Figure 4.1).

There were two major considerations when deciding on the length of this random region. First, we wanted to have a diversity compatible with GRAP display ($\approx 10^8$). Second, we wanted the random region to be semi-stable. If it was too long, we thought the transduction region may be too stable, or too far extended to allow structure-switching. If too short, there may be no suitable candidates that can optimally balance the two switch states. Drawing from known natural riboswitches, we believed that an optimal transduction region would interact strongly with the aptamer regions[97] so decided on a longer length than usual design permits (¿5bp). To balance these considerations and to incorporate some baseline stability, we chose an 18-base transduction region but inserted a G:C pair

at both ends of it. In total, we included 14 random bases into our library design (Figure 4.1 A).

4.2.2 Control Selection for ATP-Dependent Fluorescence Enhancement using MG-Biased Library

As a control, we performed a selection using GRAP display with the library described in Chapter 3 (Figure 3.1 B). We wanted to show that ATP alone cannot significantly stabilize MG binding without additional design of the library. To do so, we performed two different sorting conditions using GRAP display (Figure 2.1 Steps 3 and 4). The first type was for GRAPs that exhibited *bright* fluorescence in the *presence* of ATP (positive sort). The second type of sort was for GRAPs that exhibited *dim* fluorescence in the *absence* of ATP (negative sort). Sorting conditions were performed so as to collect only the brightest GRAPs under each condition(See Methods). In this selection, we performed three rounds of GRAP display using a positive-negative-positive sorting methodology.

Next, we performed HTS analysis using a custom MATLAB script and the Bioinformatics Toolbox to identify suitable candidates. We looked for sequences that exhibited a round-over-round increase in copy number only in the positive sort. We also looked that these sequences exhibited a round-over-round decrease in copy

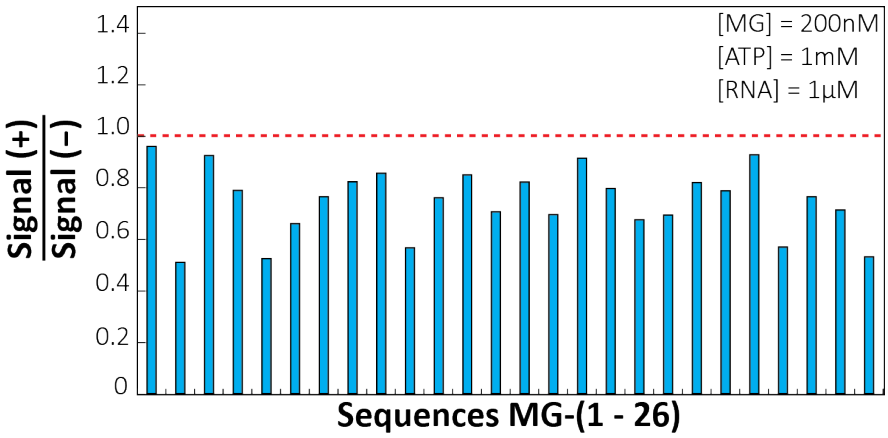

Figure 4.2: 26 sequences were identified after performing 3 rounds of GRAP display for FE aptamers in the absence and presence of ATP. No sequences exhibited the desired functionality of having a higher signal in the presence of ATP than in the absence of ATP (indicated here as a value of 1 or greater).

number when no ATP was present. In total, we identified 26 candidate sequences which were tested both in the presence and absence of ATP (See Methods).

None of the identified candidates performed the desired "turn-on" function of a fluorescent riboswitch (Figure 4.2). This suggests that performing a screen for both ATP binding and structure-switching is very likely too complex to evolve with a library diversity of only $\approx 10^8$. Having confirmed that ATP-dependent turn-on of MG is not a general phenomenon, we continued with the library design described in Section 4.2.1 (Figure 4.1).

4.2.3 Evolution of Riboswitches Through Positive and Negative Selections

In the first of two selection methods, we performed one round of GRAP display from the näive random library using the "positive" sort methodology described above. This helps enrich our pool for sequences that enhance the fluorescence of MG in the presence of ATP. However, we did not wish our pool to contain ATP-independent FE sequences, so we performed a "negative" selection as described. Thus, for a sequence to be consistently enriched it must fluoresce only in the presence of ATP. Similar to the control, we performed this selection using a positive-negative-positive selection methodology and submitted all 4 pools (Round 0, Round 1, Round 2, and Round 3) for sequencing.

Through HTS analysis (see Methods Section 4.4.5), we identified 16 sequences which exhibited consistent positive enrichment over the three rounds of selection. The identified sequences were then transcribed and serially characterized to identify those that exhibited enhanced fluorescence of MG in the presence of ATP. While 15 of the 16 sequences showed MG fluorescence in the presence of only 1 μM MG, only 4 sequences (**JY-2**, **JY-3**, **JY-7**, **JY-8**) showed increased fluorescence in the presence of 200 μM ATP (Figure 4.3). This represents significant improvement over the control selection, where all of the identified sequences showed *reduced*

fluorescence in the presence of ATP. As an example, we show sequence **JY-2** undergoing ATP-dependent enhancement in the manner desired(Figure 4.4), consistent with the expected behavior of a riboswitch. Despite this structure-switching activity, none of these sequences showed multi-fold enhancement in the presence of MG and ATP compared to MG alone.

4.2.4 Evolution of Riboswitches Through Differential Screening and HTS Analysis

In our second approach, we wished to rely more heavily on high-throughput sequencing analysis to identify sequences that enriched preferentially in the presence of ATP. Rather than relying on "positive" and "negative" sorts to enrich performing sequences, we wished to look for differences in the way that sequences performed under these conditions.

We performed one round of GRAP display from the näive random library using the "positive" sort methodology described above. In the following round, we divided our collection of GRAPs into two pools, and sorted one under "positive" sort conditions, and one under "negative" sort conditions. In both sorts, we collected 40,000 events. Because both sorts used the same pool at identical MG concentrations, we would expect that sequences which activate MG in an ATP-independent manner should enrich at similar levels. However, sequences that had

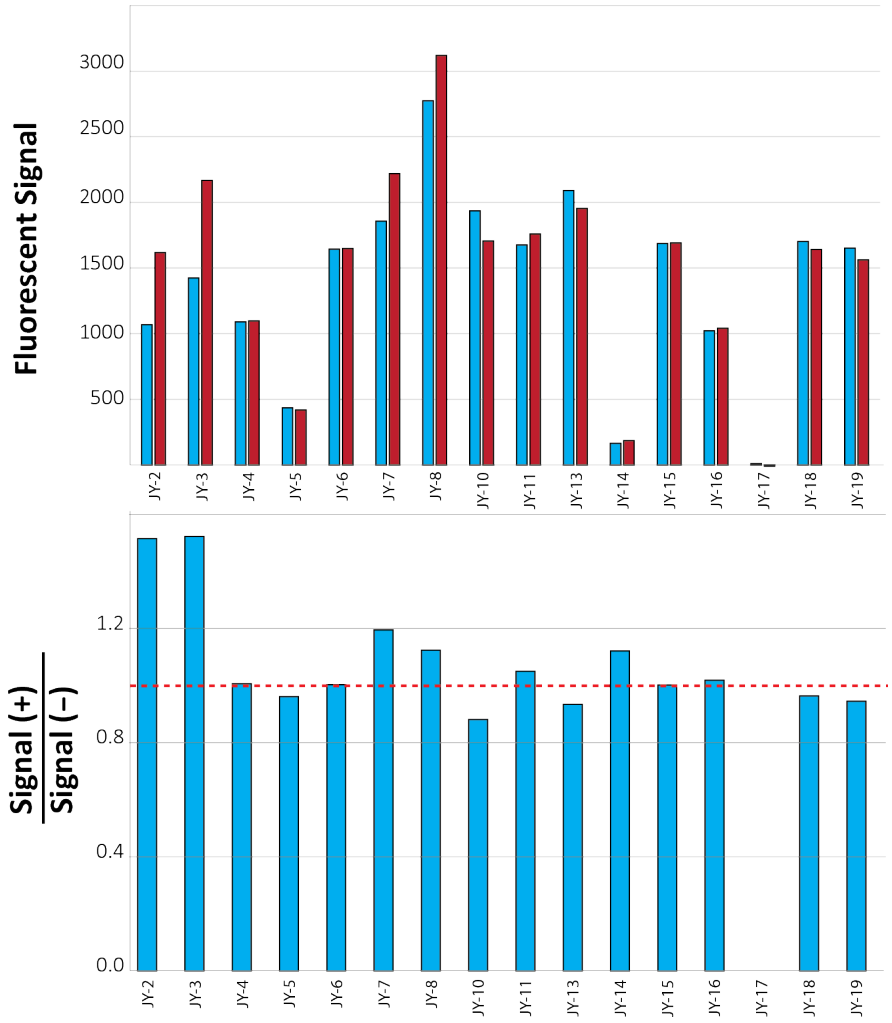

Figure 4.3: 16 sequences were identified from this selection based on enrichment in the presence or absence of ATP. **A)** The ability of these sequences to enhance the fluorescence of MG were tested in the presence and absence of ATP. 15 of these exhibited significant FE of MG. **B)** 6 sequences exhibited the desired functionality of having a higher signal in the presence of ATP than in the absence of ATP (indicated here as a value of 1 or greater). Fluorescence is in arbitrary units.

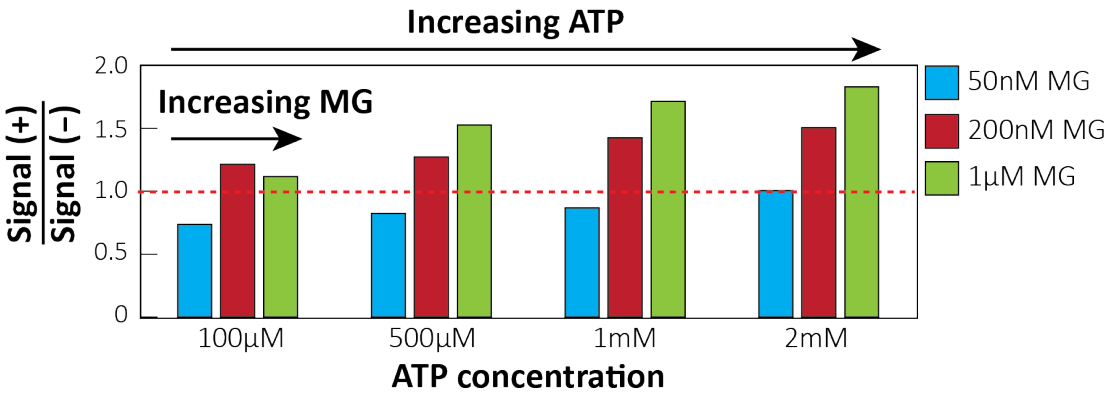

Figure 4.4: Sequence **JY-2** exhibits increased FE of MG with the addition of ATP. This is consistent with the desired switching functionality of riboswitches. RNA concentration is 1 μM.

ATP-dependent activation of MG should enrich at higher levels in the "positive" pool than in the "negative" pool. Upon sorting, the pools were not combined, and further rounds of "positive"/"negative" sorting was performed. Lastly, all pools were submitted for HTS.

Through HTS analysis, we identified 24 sequences that exhibited enrichment consistent with what would be expected for a "turn on" fluorescent riboswitch. Upon characterization (See Methods 4.4.6), we identified 19 of these sequences as yielding significant fluorescence in the presence of 1μM MG alone. Upon testing these sequences with the addition of 500 μM ATP, we identified 14 sequences which exhibited a significant fluorescence increase with the addition of ATP (Figure 4.5). One sequence, **ATP-1.3**, was found to have very little fluorescence in the absence of ATP, but many folds higher fluorescence when ATP was present. This suggests that the sequence performs a major structural reorganization upon addition of ATP. Further characterization revealed a $\approx 3-4$-fold increase in fluorescence at 1 μM MG with the addition of 200 μM ATP(Figure 4.6). This is on-par with the best reported riboswitch[126] for ATP-dependent enhancement of MG. However, the reported switch is a DNA:RNA chimera containing two ATP binding pockets which allows a greater energy contribution to structure stabilization upon ligand binding so the two can not be directly compared.

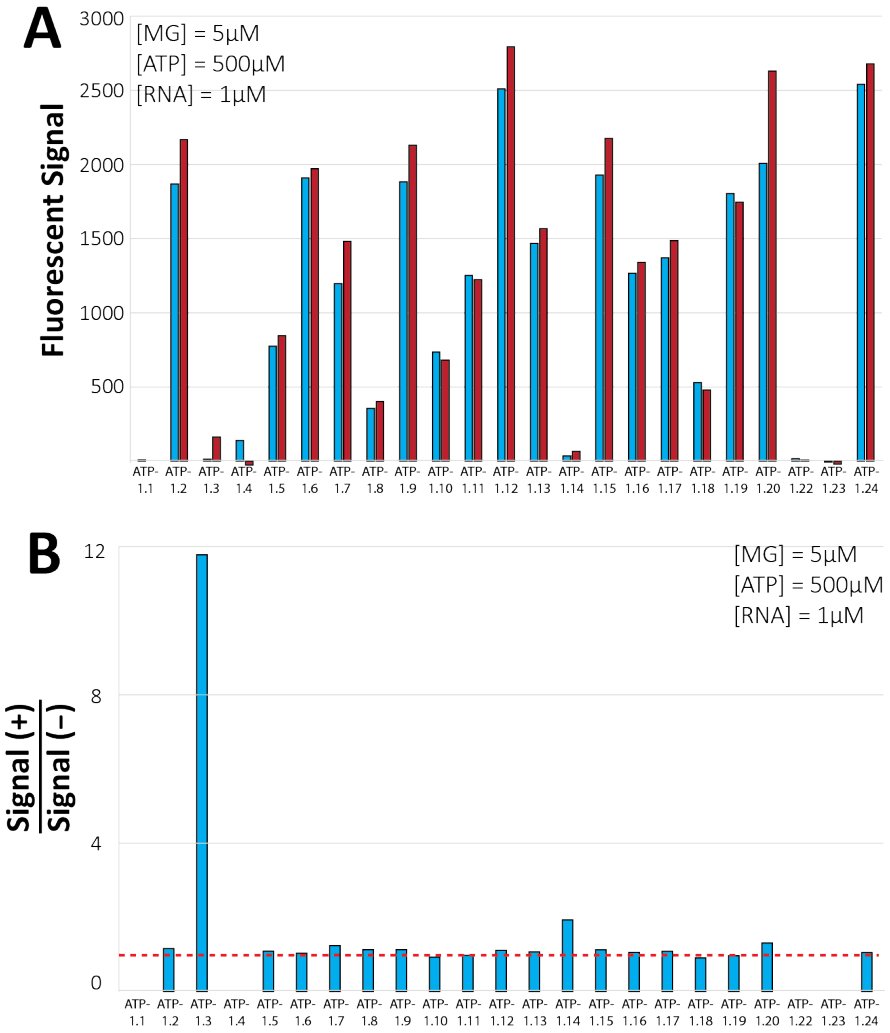

Figure 4.5: 24 sequences were identified after performing consecutive sorts using GRAP display of a single pool in the presence (positive sort) and absence (negative sort) of ATP. These sequences were positively enriched when ATP was present, and negatively enriched when ATP was absent indicating a fluorescent "turn-on" switch. **A)** The ability of these sequences to enhance the fluorescence of MG were tested in the presence and absence of ATP. **B)** 14 sequences exhibit fluorescence enhancement that is higher in the presence of ATP than without (indicated here as a value of 1 or greater), suggesting they are behaving as switches. Due to the low signal of **ATP-1.3** in the absence of ATP, we performed additional characterization to confirm the ≈12-fold increase in FE. Fluorescence is in arbitrary units.

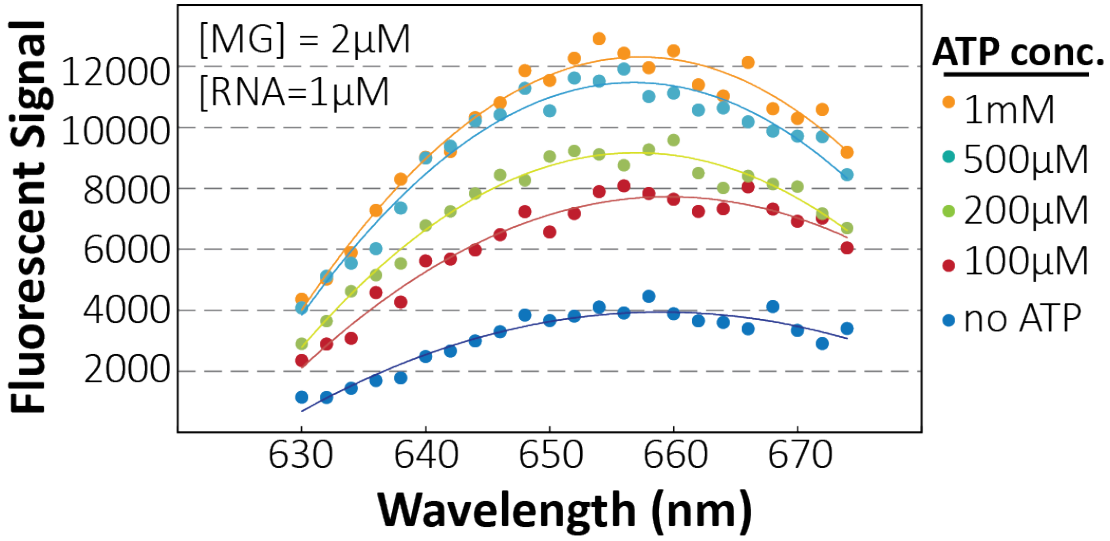

Figure 4.6: Sequence **ATP-1.3** exhibited a ≈3-fold increase in fluorescence in the presence of 500 μM ATP relative to no ATP. Measurement was performed at 2 μM MG and 1 μM RNA. Fluorescence is in arbitrary units.

4.2.5 Structural Analysis of Sequence ATP-1.3

In reported methods of rational design for riboswitches, the transduction region is designed to be unstable due to i) being short, ii) containing "weak" base pairs like A-U and G-U, or iii)having totally mismatched base pairs[101]. We wished to ascertain whether evolution could provide a method of developing more complex transduction regions (P2) that preferentially interact with the recognition (P1) and reporting (P3) regions. This mimics more natural riboswitches which undergo large structural reorganizations rather than merely going from unfolded to folded in the presence of their ligand[95,97,188].

We can show GRAP display allows this by comparing the predicted two-state folding of sequence **ATP-1.3** (Figure 4.7). Mfold software predicts a highly stable inactive structure with the P2 region interacting strongly with P1. Indeed, the active structure is predicted not to fold as designed due to strong destabilization in the P2 region which prevents the G:C clamps on either end of the transduction region from correctly folding. The predicted energy difference of the structures is nearly $9\frac{kcal}{mol}$. This suggests the MG and ATP provide significant stabilization to promote folding of the active structure. For reference, many natural riboswitches exhibit a ΔG for small molecule-RNA interactions from $7 - 12\frac{kcal}{mol}$ [97].

While these are only predicted free energies and do not take into account environmental affects or any of the numerous other potential folding states, we see that the large RNA:RNA interactions present in **ATP-1.3** appear to mimic more natural riboswitches which incorporate many interactions between the recognition and active region of the switch. This differs strongly from the mostly unstructured regions that have been rationally designed [101–103,126,161].

With certainty, we can say that usual methods of rational design would fail to predict a transduction region such as that in **ATP-1.3**. This is because of 1) the long (9bp) length of the P2 region, 2) the strong interactions predicted between the P2 and P3 regions, and 3) the huge stability difference predicted between the active and inactive state. Indeed, as we perfect the methods used in this selection, we believe GRAP display may allow discovery of numerous transduction regions that would otherwise be difficult or impossible to predict. This could lead to great advances into the rational design and *in vivo* or *in vitro* application of riboswitches.

4.3 Conclusions

Riboswitches are of vast importance in nature where species utilize target-dependent switching of RNA for functions as diverse as mRNA silencing, cleavage, splicing, and protein recruitment [95]. However, current methods of riboswitch development rely on altering known riboswitch structures in an effort to impart additional

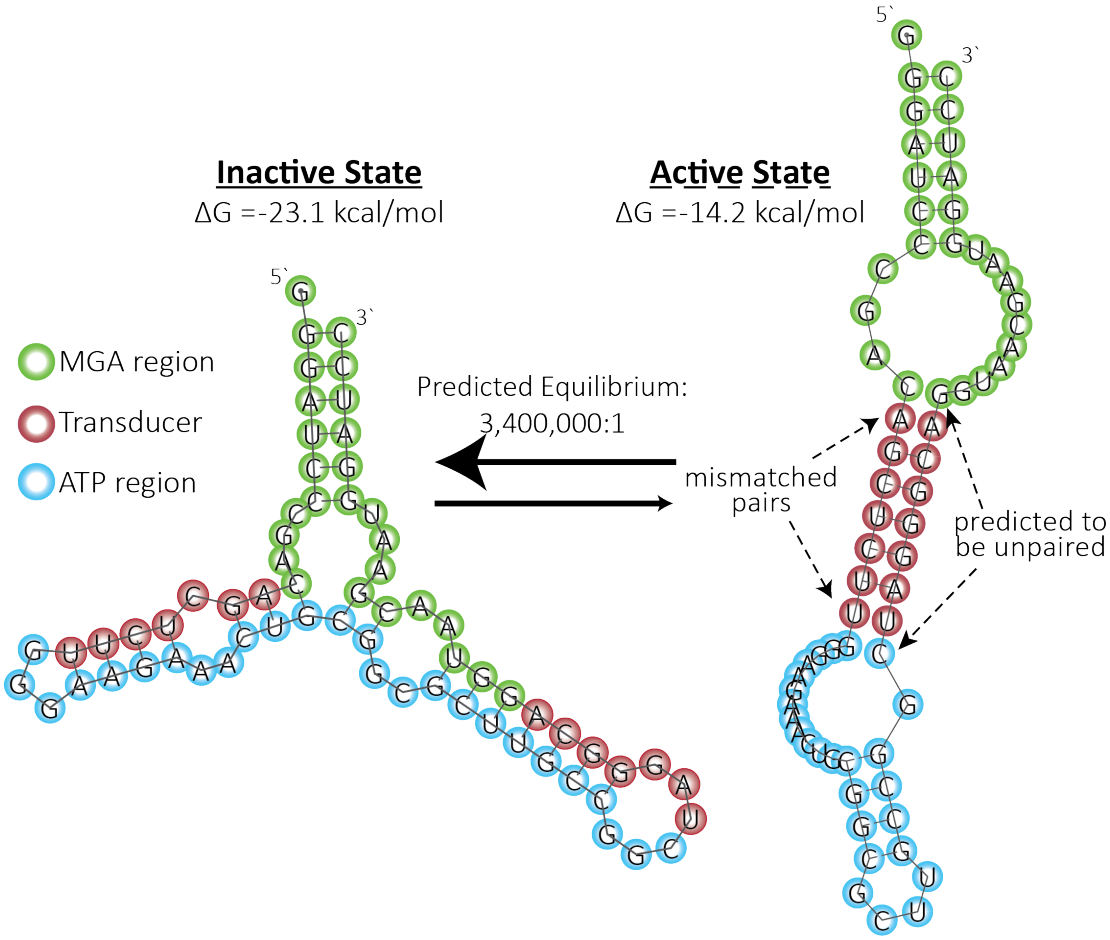

Figure 4.7: mfold structure predicted for sequence **ATP-1.3**. The inactive state (misfolded) is predicted to be highly structured and stable with a predicted free energy of $-23.1 kcal/mol$. The correctly folded state is highly destabilized by the incorporation of mismatched pairs on either end of the transduction region. The predicted free energy for this structure is $-14.2 kcal/mol$. This results in a predicted 3,400,000:1 distribution of the aptamer in the inactive state compared to the active state.

function, or laborious *in vivo* selections for transduction regions. While both methods have allowed discovery of functional riboswitches for various targets, neither allows both a thorough analysis of sequence space and easy adaptation to new targets.

For this reason, we applied GRAP display to search for target-dependent activation of a fluorescence enhancing RNA. We chose two approaches to address this problem. In the first, we performed consecutive rounds of "positive" and "negative" selection against ATP. In the second, we performed simultaneous rounds of "positive" and "negative" selection using a single pool, and analyzed HTS data to identify potential candidates. Both methods produced numerous aptamers with ATP-dependent fluorescence enhancement indicative of switching. One sequence, **ATP-1.3**, provided a strong enhancement of MG only in the presence of ATP - exhibiting performance rivaling the best-designed MG/ATP switch to date[126]. Additionally, the transduction region of this sequence does not appear to follow the rules usually applied in the rational design of riboswitches. Instead, the predicted free energies of folding and strong overlap of the transduction region with the aptamer region more closely mimics the features of natural riboswitches. This suggests that GRAP display may be an effective manner of evolving switches with complexity beyond what rational design allows.

While there remains much to explore in regards to the optimal design of an evolve-able riboswitch library, we believe these results show that GRAP display is a promising method of addressing the need for new riboswitches. Not only can evolution discover switches with complex interactions between the different regions of a switch, but these interactions more closely mimic natural riboswitches. Additionally, this manner of library design can be easily applied to any target with a known aptamer by simply switching out one region of the library with a new motif and performing 2-3 rounds of GRAP display. Future selections should yield a number of insights into modular riboswitch discovery towards the goal of arbitrary cellular computing and control, as well as development of effective *in vitro* switches for use as biosensors.

4.4 Methods

GRAP display was prepared and performed in the manner previously described (Section 2.4). Buffer recipes and reagents remain as previously described. Additionally, sequences were ordered, synthesized, purified, and characterized as described previously.

4.4.1 DNA preparation

The single-stranded DNA library and all primers used were synthesized by Integrated DNA Technologies. Modified oligos were purified using HPLC, while all others underwent standard desalting. Randomized bases were hand-mixed at a 25:25:25:25 ratio. The DNA library consisted of molecules containing a conserved ATP motif flanked by randomized 7nt regions which were in turn flanked by a conserved MG motif and further by PCR primer-binding sequences (Figure 4.1).

Identical forward primers were used as described previously (Section 2.4). For the selections performed in this chapter (except for the control selection), we redesigned the RP in order to prevent cross-contamination from previous selections. Oligos used in this selection are as follows:

Oligo	**Sequence** ($5' \rightarrow 3'$)
Library	TAATACGACT CACTATAGGG ACACAATGGA CGGGGATCCC GACNNNNNNN GGGAAGAAAC TGCGGCGCTT GCCGGCNNNN NNNGGTAACG AATGGATCC
FP	TAATACGACT CACTATAGGG ACACAATGGA CG
RP	GGATCCATTC GTTACC
RPpT	TTTTTTTTTT TTTTTTTTTT TTTTTGGATC CATTCGTTAC C
FAM-RPpT	/56-FAM/-TTTTTTTTTT TTTTTTTTTT TTTTTGGATC CATTCGTTAC C
AmM-FP	/5AmMC6//iSp18//iSp18//iSp18/ /iSp18/-TAATACGACT CACTATAGGG ACACAATGGA CG
AmM-pT	/5AmMC6//iSp18//iSp18//iSp18/ /iSp18/-TTTTTTTTTT TTTTTTTTTT TTTTT
FP-647	/5Alex647N/-TAATACGACT CACTATAGGG ACACAATGGA CG
FPc-FAM	/56-FAM/-CGTCCATTGT GTCCCTATAG TGAGTCGTAT TA
pA-FAM	/56-FAM/-AAAAAAAAAA AAAAAAAAAA AAAAA

4.4.2 Control selection

In the control selection, we used the library from Chapter 3 and all oligos as described there. We performed two rounds of pre-enrichment SELEX using the methods described in Section 2.4 with ATP immobilized on agarose beads (Sigma, final concentration ≈2 mM ATP). In Round 1 of pre-SELEX, we eluted the RNA using free 10 mM ATP (Thermo Fischer Scientific) and 1 μM MG instead of heat.

In Round 2, we used fewer ATP-agarose beads for an estimated concentration of 200 μM ATP.

We performed the initial (Round 0) GRAP display sort at 1 μM MG and 1 mM ATP, collecting around 0.2% of the positive population. In Round 1, we reduced the concentration of MG to 500 nM and collected $\approx$ 0.1% of particles. In the negative Round 2 sort, we collected a small fraction of non-fluorescent particles at 50 nM MG. In the final positive Round 3 sort, we collected particles that exhibited a fluorescent signal at 5 nM MG and 100 μM ATP.

HTS was performed on all pools as described previously, and 24 sequences were identified that exhibited positive enrichment in the presence of ATP during one or more rounds. None exhibited a strong enrichment *only* in the presence of ATP, which is what we would expect if no functional switches are present in the pool.

Upon characterizing the 24 sequences, we found that none exhibited positive signal gain upon addition of ATP. We confirmed thsi by preparing Round 1 GRAPs and incubating them with FAM-labeled ATP. We found that these displayed very little signal which suggests that we did not have ATP-binding aptamers in the Round 1 pool.

4.4.3 Positive and negative selection

We prepared a Round 0 GRAP pool using the new library design. We performed one round of GRAP display as pre-enrichment, without any pre-SELEX, and collected fluorescent GRAPs at 1 μM MG and 100 μM ATP. These target concentrations were used for all further rounds of selection. We collected $\approx 0.1\%$ of the GRAP population. In all rounds of this selection, we collected ≈30,000 events.

From the Round 1 pool, we collected 0.7% of particles in the presence of ATP (positive sort). With the Round 2 pool, we collected an arbitrary fraction of the GRAPs that expressed DNA and did *NOT* exhibit a strong MG signal (negative sort). We sorted the Round 3 pool in the presence of ATP and collected $\approx 0.1\%$ of the brightest GRAPs. All pools were then subjected to HTS

4.4.4 Differential screening selection

In our second selection attempt, we used the same DNA that was prepared from sorting the Round 0 pool in the first selection attempt. We then prepared a new aliquot of Round 1 GRAPs.

Round 1 GRAPs ($\approx 10^8$ particles) were resuspended in 500 μL TNaTE, and physically separated into two tubes. The beads were concentrated using the MPC-S and resuspended into 5mL of PBSMT buffer containing 1 μM MG, each. ATP was

added to one tube to a final concentration of 100 μM. Both of these tubes were incubated concurrently, and sorted individually so that the top 0.1% of fluorescent particles were collected. Approximately 40,000 events were collected from each pool to yield the R2P (from the positive sort) and R2N (from the negative sort) pools. For the next round of GRAP display under the same target conditions, we collected $\approx$ 31,000 events to yield the R3PP, R3PN, R3NP, and R3NN pools. GRAPs were again prepared and sorted, whereby $\approx$21,000 events were collected to yield the R3PPP, R3PPN, R3PNP, R3PNN, R3NPP, R3NPN, R3NNP, R3NNN pools. A total of 15 pools were then subjected to HTS.

4.4.5 Sequencing analysis

HTS data was analyzed using the methods described previously. We looked for sequences that exhibited positive enrichment only when sorted in the presence of ATP. Though none followed a clear trend of positive and negative enrichment as expected, we were able to identify 26 sequences from the control selection, 16 from the first selection method(positive-negative-positive), and 24 from the second selection method(differential sequence analysis).

A control sequence (GGGATCCCGA CCTGGGGTGG GAAGAAACTG CGGCGCTTGC CGGCACCCCA GGGTAACGAA TGGATCC), which was fully paired in the transduction region of the library, was ordered in order to test

the folding of the library on beads. We neglected to make this sequence differ enough from the library to assure that it would not be PCR amplified and come to enrich in our pools. Our sequencing data revealed that our pools were $60-80\%$ contaminated by this control sequence, which highly limited our ability to perform accurate and in-depth sequencing analysis. Indeed, it is a testament to GRAP display that this control sequence was able to enter our pool through minimal cross-contamination from shared equipment despite our best efforts to prevent this. Nonetheless, the identified sequences from the second selection method which yielded **ATP-1.3** are presented in Table 4.1.

4.4.6 Sequence characterization

All sequences from the three discussed selection were ordered from the Protein And Nucleic acid (PAN) Facility at Stanford University. They were amplified, transcribed, and purified as described previously. In general, sequences were tested at one concentration of MG in the presence and absence of ATP. This was done in order to determine if any sequences exhibited differential functionality. In the control selection, MG concentration for these initial tests was set at 200 nM and ATP at 1 mM with RNA at 1 μM (Figure 4.2). For selection methodologies 1 and 2, MG was set at 5 μM MG, ATP at 500 μM, and RNA at 1 μM(Figures 4.3 and

Oligo	**Sequence** ($5' \rightarrow 3'$)
ATP.1.1	AAGCCCCGGGAAGAAACTGCGGCGCGTGCCGGCAATGTTA
ATP.1.2	TTGGTATGGGAAGAAACTGCGGCGCTTACCGGCATACCGA
ATP.1.3	AGCTCTTGGGAAGAAACTGCGGCGCTTGCCGGCTAGGGCA
ATP.1.4	GAAACTGGGGAAGAAACTGCGGCGCTTGCCGGCGCACGTG
ATP.1.5	TCACGTAGGGAAGAAACTGCGGCGCTTGCCGGCAACAGGA
ATP.1.6	TCTATCCGGGAAGAAACTGCGGCGCTTGCCGGCTTAATGA
ATP.1.7	TCACTGCGGGAAGAAACTGCGGCGCTTGCCGGCGGCTGGA
ATP.1.8	CGCGTGTGGGAAGAAACTGCGGCGCTTGCCGGCCCACACA
ATP.1.9	ACAACTCCGGAAGAAACTGCGGCGCTTGCCGGCGAGGTGA
ATP.1.10	ACCCTCAGGGAAGAAACTGCGTCGCTTGCCGGCCGTGGGT
ATP.1.11	AATTTTAGGGAAGAAACTGCAGCGCTTGCCGGCTTAAATT
ATP.1.12	ATCTTGCGGGGAGAAACTGCGGCGCTTGCCGGCATTAGGT
ATP.1.13	TTTGATCGGGAAGAAACTGCGGCGCTTGCCGGTGATCAGA
ATP.1.14	TCACTGCGGGAAGAAACTGCGGCGCTTGCCGGCATATTGA
ATP.1.15	ACCGTCGGGGAAGAAACTGCGGCGCTTGCCGGCAATAAGA
ATP.1.16	AGTGTGCGGGAAGAAACTGCGGCGCTTGCCGGCTTACACT
ATP.1.17	TCTACGCGGGAAGAAACTGCGGCGCTTGCAGGCATAAAGA
ATP.1.18	TGACACCGGGGAGAAACTGCGGCGCTTGCCGGCAAAATCA
ATP.1.19	AAGTATTGGGAAGAAACTGCAGCGCTTGCCGGCTTCACTT
ATP.1.20	TAGATCCGGGAAGAAACTGCGGCGCTTGCCGGCCAAACTA
ATP.1.21	TCGCGTGGGGAAGAAACTGCGTCGCTTGCCGGCCCAGCGA
ATP.1.22	GTAATAAGGGAAGAAACTGCGGCGCTTGCCGGCAAAGATC
ATP.1.23	TAAGCACGGGAAGAAACTGCGGCGCTTGCCGGCAACAACT
ATP.1.24	TGCGTCAGGGAAGAAACTGCGGCGCTTGCCGGCCCGCGCA

Table 4.1: Sequencing results from the selection method relying on differential screening and sequence analysis. Shown is *only* the region containing the transducing region and ATP motif.

4.5). The concentrations used for these characterizations were $\approx 5 - fold$ above the selection conditions used.

Once candidate sequences were identified, they were further tested to confirm activity. Sequence **JY-2** (Figure 4.4) was prepared in duplicate at a final concentration of 1 μM RNA in 0 nM MG, 50 nM MG, 200 nM MG, and 1 μM MG. Then, ATP was added to one well at each MG concentration and the fluorescence signal was measured. This process was repeated for concentrations of 100 μM, 500 μM, 1 mM, and 2 mM ATP and the results reported.

Sequence **ATP-1.3** (Figure 4.6) was prepared in PBSMT, in triplicate, at a concentration of 1 μM RNA at ATP concentrations of 0 μM, 100 μM ATP, 200 μM ATP, 500 μM ATP, and 1 mM ATP. These samples were then incubated at concentrations of 500 nM MG, 2 μM MG, and 5 μM MG and fluorescence measured using a Tecan M1000 plate reader.

Chapter 5

Conclusion

5.1 Summary

Over the last decade, we have learned that RNA is a much more complex and widely applied molecule than originally believed. Rather than being a simple intermediate between DNA and protein, RNA serves a number of diverse functions in the cell ranging from regulating gene expression to environmental sensing. However, many of our directed evolution approaches for developing new functional RNA rely on enriching sequences based simply on their ability to recognize a target of interest. Towards expanding the capabilities of directed evolution to discover complex RNA function, my dissertation demonstrates the unique capabilities of a new approach

to *in vitro* RNA selection, termed Gene-linked RNA Aptamer Particle (GRAP) display, and describes multiple applications where it can be used to develop novel, highly functional RNA aptamers.

In Chapter 2, we introduced GRAP display - an *in vitro* selection method capable of enriching RNA sequences based on their interactions with a fluorescent target. We demonstrate this capability by performing a selection against the non-fluorescent small molecule dye, malachite green (MG). Upon binding MG, an RNA aptamer can enhance the fluorescence in a manner independent of binding affinity. As such, we were able to demonstrate that GRAP display successfully identified sequences with desirable fluorescent properties that would otherwise be missed using SELEX.

We isolated aptamers with affinities ranging from 20 nM to >2 μM, and resulting in a malachite green quantum yield in excess of 3,800-fold above background. Additionally, because of the unique manner by which GRAP display measures the fluorescence output of every individual aptamer species, we showed that we can intentionally isolate aptamers with distinct spectral properties upon interaction with a single chromophore. For example, one group of sequences shared a short, 15nt motif that yielded an emission profile for malachite green that was significantly red-shifted from the known sequence and had an emission maximum of 675nm.

Lastly, we performed numerous characterization of the discovered aptamers and demonstrated their utility in *in vivo* environments.

Next, we applied GRAP display in Chapter 3 for the re-evolution of a known FE motif in order to improve its function. We showed that, by incorporating additional structure to the sequence, we could improve on both the affinity and brightness of the known aptamer. We also saw that one sequence, through incorporation of a single nucleotide mismatch from other characterized sequences(and in the conserved motif, no less) exhibited a greatly reduced affinity and increased quantum yield. The location and effect of this mismatch suggests that the previously-solved crystal structure appears to show a unique tetraloop folding motif which is in agreement with the crystal structure but contrary to the views of the original authors.

With the understanding that known library mutagenesis techniques exhibit massive redundancy that makes them unsuitable for particle-based methods, we discussed design parameters for future random libraries. Here, we took the known method of mutagenized aptamer library design and suggested an optimization step in the design to reduce redundancy in highly-conserved (having a small number of mutations) sequences. In order to maximize the amount of sequence space capable of being explored, while also improving the efficiency of selection, we broadly suggest an increased mutation rate that provides improved sequence diversity for the non-redundant exploration of sequence space around a known motif.

With the goal of eventually discovering RNA aptamers exhibiting highly complex and coordinated functions, we next applied GRAP display for the development of a simple fluorescence-reporting riboswitch which functions by transducing a target recognition event into a fluorescence output. We designed a random library that incorporates known motifs for the discovery of an ATP-dependent fluorescence-enhancing aptamer.

We wished to show that evolution could construct riboswitches that go beyond the simple design parameters commonly used for riboswitch engineering and instead produce regions that more closely mimic natural riboswitches. So, we discussed and performed two new selection methodologies and showed that both are capable of isolating ATP-dependent riboswitches. One method yielded Sequence **ATP-1.3** which exhibits multi-fold FE in the presence of ATP and had a non-conventional transduction region that does not follow the parameters usually followed during rational design. This shows that GRAP display is capable of evolving entirely new riboswitch domains which no other *in vitro* selection methodology can perform.

RNA is an incredibly functional biomolecule, and particle-based display methods have repeatedly demonstrated their unparalleled capacity for reliably and effectively isolating highly functional nucleic acid aptamers. As the first RNA-based particle display selection methodology, we believe that GRAP display is a profound step forward towards realizing the ultimate potential and practical uses of functional

RNA. While I will leave the numerous potential applications of this technique to the imagination of the reader, I sincerely hope GRAP display can be further applied towards the evolution of complex RNA machinery capable of more than simply binding to a target molecule.

Bibliography

[1] Osamu Shimomura, Frank H. Johnson, and Yo Saiga. *J. Cell. Comp. Physiol.*, 59(3):223–239, 1962. doi: 10.1002/jcp.1030590302.

[2] Satoshi Inouye and Frederick I. Tsuji. *FEBS Lett.*, 341(2-3):277–280, 1994. doi: 10.1016/0014-5793(94)80472-9.

[3] Martin Chalfie, Yuan Tu, Ghia Euskirchen, William W Ward, and Douglas C Prashert. *Science*, 263(5148):802–805, 1994.

[4] Won-Ki Huh, James V. Falvo, Luke C. Gerke, Adam S. Carroll, Russell W. Howson, Jonathan S. Weissman, and Erin K. O'Shea. *Nature*, 425(6959): 686–691, 2003. doi: 10.1038/nature02026.

[5] B. N. G. Giepmans. *Science*, 312(5771):217–224, apr 2006. doi: 10.1126/science.1124618.

[6] Michael B Elowitz and Stanislas Leibler. *Nature*, 403(6767):335–338, jan 2000. doi: 10.1038/35002125.

[7] Chang Deng Hu, Yurii Chinenov, and Tom K. Kerppola. *Mol. Cell*, 9(4): 789–798, 2002. doi: 10.1016/S1097-2765(02)00496-3.

[8] W. E. Moerner. *Science*, 283(5408):1670–1676, 1999. doi: 10.1126/science.283.5408.1670.

[9] Roger Y Tsien. *Ann. Rev. Biochem.*, 67:509–544, 1998.

[10] Marc Zimmer. *Chem. Rev.*, 102(3):759–781, 2002. doi: 10.1021/cr010142r.

[11] Nathan C Shaner, George H Patterson, and Michael W Davidson. *J. Cell Sci.*, 120(24):4247–4260, dec 2007. doi: 10.1242/jcs.005801.

[12] Koji Nakanishi, Nobuo Furutachi, Makoto Funamizu, Dezider Grunberger, and I Bernard Weinstein. *J. Am. Chem. Soc.*, 92(26):7617–9, 1970.

[13] Elena V. Dolgosheina and Peter J. Unrau. *Wiley Interdiscip. Rev. RNA*, 7 (6):843–851, 2016. doi: 10.1002/wrna.1383.

[14] Mark Muyskens. *J. Chem. Educ.*, 83(5):765, 2006. doi: 10.1021/ed083p765.

[15] John Frederick William Herschel. *Philos. Trans. R. Soc. London*, 135: 143–145, 1845.

[16] Joseph R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Springer US, Boston, MA, 1983. doi: 10.1007/978-1-4615-7658-7.

[17] Mark A. Haidekker and Emmanuel A. Theodorakis. *Org. Biomol. Chem.*, 5 (11):1669–1678, 2007. doi: 10.1039/B618415D.

[18] Zujin Zhao, Ping Lu, Jacky W. Y. Lam, Zhiming Wang, Carrie Y. K. Chan, Herman H. Y. Sung, Ian D. Williams, Yuguang Ma, and Ben Zhong Tang. *Chem. Sci.*, 2(4):672, 2011. doi: 10.1039/c0sc00521e.

[19] Yuning Hong, Jacky W. Y. Lam, and Ben Zhong Tang. *Chem. Soc. Rev.*, 40 (11):5361, 2011. doi: 10.1039/c1cs15113d.

[20] Anton Simeonov. *Science*, 290(5490):307–313, oct 2000. doi: 10.1126/science.290.5490.307.

[21] Michael J. Taussig, Oda Stoevesandt, Carl A. K. Borrebaeck, Andrew R. Bradbury, Dolores Cahill, Christian Cambillau, Antoine de Daruvar, Stefan Dübel, Jutta Eichler, Ronald Frank, Toby J. Gibson, David Gloriam, Larry Gold, Friedrich W. Herberg, Henning Hermjakob, Jörg D. Hoheisel, Thomas O. Joos, Olli Kallioniemi, Manfred Koegl, Zoltán Konthur, Bernhard Korn, Elisabeth Kremmer, Sylvia Krobitsch, Ulf Landegren, Silvère van der Maarel, John McCafferty, Serge Muyldermans, Per-Åke Nygren, Sandrine Palcy, Andreas Plückthun, Bojan Polic, Michael Przybylski, Petri Saviranta, Alan Sawyer, David J. Sherman, Arne Skerra, Markus Templin, Marius Ueffing, and Mathias Uhlén. *Nat. Methods*, 4(1):13–17, 2007. doi: 10.1038/nmeth0207-187a.

[22] Karen Colwill and Susanne Gräslund. *Nat. Methods*, 8(7):551–558, 2011. doi: 10.1038/nmeth.1607.

[23] Sachdev S. Sidhu. *FEBS Lett.*, 586(17):2778–2779, 2012. doi: 10.1016/j.febslet.2012.05.044.

[24] Marco Mascini. *Anal. Bioanal. Chem.*, 390(4):987–988, 2008. doi: 10.1007/s00216-007-1769-y.

[25] Anthony D Keefe, Supriya Pai, and Andrew Ellington. *Nat. Rev. Drug Discov.*, 9(7):537–550, 2010. doi: 10.1038/nrd3249.

[26] Jennifer Bordeaux, Allison W. Welsh, Seema Agarwal, Elizabeth Killiam, Maria T. Baquero, Jason A. Hanna, Valsamo K. Anagnostou, and David L. Rimm. *Biotechniques*, 48(3):197–209, 2010. doi: 10.2144/000113382.

[27] Andrew Bradbury and Andreas Plückthun. *Nature*, 518(7537):27–29, feb 2015. doi: 10.1038/518027a.

[28] Matthew A Helsby, Joe R Fenn, and Andrew D Chalmers. *F1000Research*, 2(3):153, 2013. doi: 10.12688/f1000research.2-153.v2.

[29] Sumedha D. Jayasena. *Clin. Chem.*, 45(9):1628–1650, 1999.

[30] E N Brody and L Gold. *J. Biotechnol.*, 74(1):5–13, 2000. doi: 10.1016/S1389-0352(99)00004-5.

[31] Shahid M Nimjee, Christopher P Rusconi, and Bruce a Sullenger. *Annu. Rev. Med.*, 56(FEBRUARY):555–83, 2005. doi: 10.1146/annurev.med.56.062904.144915.

[32] Saw Y i Toh, Marimuthu Citartan, Subash C B Gopinath, and Thean Hock Tang. *Biosens. Bioelectron.*, 64:392–403, 2015. doi: 10.1016/j.bios.2014.09.026.

[33] Geoffrey S. Baird. *Am. J. Clin. Pathol.*, 134:529–531, 2010. doi: 10.1309/AJCPFU4CG2WGJJKS.

[34] Teresa Mairal, Veli Cengiz Özalp, Pablo Lozano Sánchez, Mònica Mir, Ioanis Katakis, and Ciara K. O'Sullivan. *Anal. Bioanal. Chem.*, 390(4):989–1007, 2008. doi: 10.1007/s00216-007-1346-4.

[35] Regina Stoltenburg, Christine Reinemann, and Beate Strehlitz. *Biomol. Eng.*, 24(4):381–403, 2007. doi: 10.1016/j.bioeng.2007.06.001.

[36] Renee K. Mosing, Shaun D. Mendonsa, and Michael T. Bowser. *Anal. Chem.*, 77(19):6107–6112, 2005. doi: 10.1021/ac050836q.

[37] Oliver Kensch, Bernard A. Connolly, Heinz Jürgen Steinhoff, Alistair McGregor, Roger S. Goody, and Tobias Restle. *J. Biol. Chem.*, 275(24):18271–18278, 2000. doi: 10.1074/jbc.M001309200.

[38] R D Jenison, S C Gill, A Pardi, and B Polisky. *Science*, 263(5152):1425–9, 1994. doi: 10.1126/science.7510417.

[39] Albert Geiger, Petra Burgstaller, Herbert Von der Eltz, Albert Roeder, and Michael Famulok. *Nucleic Acids Res.*, 24(6):1029–1036, 1996. doi: 10.1093/nar/24.6.1029.

[40] L C Bock, L C Griffin, J a Latham, E H Vermaas, and J J Toole. *Nature*, 355(6360):564–566, 1992. doi: 10.1038/355564a0.

[41] D E Huizenga and J W Szostak. *Biochemistry*, 34(2):656–665, 1995. doi: 10.1021/bi00002a033.

[42] Dion a Daniels, Hang Chen, Brian J Hicke, Kristine M Swiderek, and Larry Gold. *Proc. Natl. Acad. Sci. U. S. A.*, 100(26):15416–15421, 2003. doi: 10.1073/pnas.2136683100.

[43] Joshua K. Herr, Joshua E. Smith, Colin D. Medley, Dihua Shangguan, and Weihong Tan. *Anal. Chem.*, 78(9):2918–2924, 2006. doi: 10.1021/ac052015r.

[44] Junji Kawakami, Hirofumi Imanaka, Yukie Yokota, and Naoki Sugimoto. *J. Inorg. Biochem.*, 82(1-4):197–206, 2000. doi: 10.1016/S0162-0134(00)00158-6.

[45] Subash C B Gopinath, Tomoko S. Misono, Kazunori Kawasaki, Takafumi Mizuno, Masaki Imai, Takato Odagiri, and Penmetcha K R Kumar. *J. Gen. Virol.*, 87(3):479–487, 2006. doi: 10.1099/vir.0.81508-0.

[46] a D Ellington and J W Szostak. *Nature*, 346(6287):818–22, 1990. doi: 10.1038/346818a0.

[47] C Tuerk and L Gold. *Science*, 249(4968):505–510, 1990. doi: 10.1126/science.2200121.

[48] David H J Bunka, Olga Platonova, and Peter G. Stockley. *Curr. Opin. Pharmacol.*, 10(5):557–562, 2010. doi: 10.1016/j.coph.2010.06.009.

[49] Eun Jeong Cho, Joo-Woon Lee, and Andrew D Ellington. *Annu. Rev. Anal. Chem. (Palo Alto. Calif).*, 2:241–64, 2009. doi: 10.1146/annurev.anchem.1.031207.112851.

[50] R. E. Wang, H. Wu, Y. Niu, and J. Cai. *Curr. Med. Chem.*, 18(27):4126–4138, 2011. doi: 10.2174/092986711797189565.

[51] Sulay D. Jhaveri, Romy Kirby, Rick Conrad, Emily J. Maglott, Michael Bowser, Robert T. Kennedy, Gary Glick, and Andrew D. Ellington. *J. Am. Chem. Soc.*, 122(11):2469–2473, 2000. doi: 10.1021/ja992393b.

[52] Manjula Rajendran and Andrew D. Ellington. *Nucleic Acids Res.*, 31(19): 5700–5713, 2003. doi: 10.1093/nar/gkg764.

[53] Razvan Nutiu and Yingfu Li. *Chem. - A Eur. J.*, 10(8):1868–1876, 2004. doi: 10.1002/chem.200305470.

[54] Olaf Heidenreich and F Eckstein. *J. Biol. Chem.*, 267(3):1904–9, jan 1992.

[55] Jeremy R Babendure, Stephen R Adams, and Roger Y Tsien. *J. Am. Chem. Soc.*, 125(48):14716–14717, dec 2003. doi: 10.1021/ja037994o.

[56] W Kusser. *J. Biotechnol.*, 74(1):27–38, 2000. doi: 10.1016/S1389-0352(99) 00002-1.

[57] Monya Baker. *Nature*, 521(7552):274–276, 2015. doi: 10.1038/521274a.

[58] Michael M K Vu, Nora E. Jameson, Stuart J. Masuda, Dana Lin, Rosa Larralde-Ridaura, and Andrej Lupták. *Chem. Biol.*, 19(10):1247–1254, 2012. doi: 10.1016/j.chembiol.2012.08.010.

[59] Yuki Fujimoto, Yoshikazu Nakamura, and Shoji Ohuchi. *Biochimie*, 94(9): 1900–1909, 2012. doi: 10.1016/j.biochi.2012.05.003.

[60] Larry Gold, Deborah Ayers, Jennifer Bertino, Christopher Bock, Ashley Bock, Edward N. Brody, Jeff Carter, Andrew B. Dalby, Bruce E. Eaton, Tim Fitzwater, Dylan Flather, Ashley Forbes, Trudi Foreman, Cate Fowler, Bharat Gawande, Meredith Goss, Magda Gunn, Shashi Gupta, Dennis Halladay, Jim Heil, Joe Heilig, Brian Hicke, Gregory Husar, Nebojsa Janjic, Thale Jarvis, Susan Jennings, Evaldas Katilius, Tracy R. Keeney, Nancy Kim, Tad H. Koch, Stephan Kraemer, Luke Kroiss, Ngan Le, Daniel Levine, Wes Lindsey, Bridget Lollo, Wes Mayfield, Mike Mehan, Robert Mehler, Sally K. Nelson, Michele Nelson, Dan Nieuwlandt, Malti Nikrad, Urs Ochsner, Rachel M. Ostroff, Matt Otis, Thomas Parker, Steve Pietrasiewicz, Daniel I. Resnicow, John Rohloff, Glenn Sanders, Sarah Sattin, Daniel Schneider, Britta Singer, Martin Stanton, Alana Sterkel, Alex Stewart, Suzanne Stratford, Jonathan D. Vaught, Mike Vrkljan, Jeffrey J. Walker, Mike Watrobka, Sheela Waugh, Allison Weiss, Sheri K. Wilcox, Alexey Wolfson, Steven K. Wolk, Chi Zhang, and Dom Zichi. *PLoS One*, 5(12), 2010. doi: 10.1371/journal.pone.0015004.

[61] Sven Klussmann, Alexis Nolte, Rolf Bald, Volker A. Erdmann, and Jens P. Fürste. *Nat. Biotechnol.*, 14(9):1112–1115, 1996. doi: 10.1038/nbt0996-1112.

[62] Anthony D. Keefe and Sharon T. Cload. *Curr. Opin. Chem. Biol.*, 12(4): 448–456, aug 2008. doi: 10.1016/j.cbpa.2008.06.028.

[63] V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S.-Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn, and P. Holliger. *Science*, 336(6079):341–344, apr 2012. doi: 10.1126/science.1217622.

[64] Abdullah Ozer, John M Pagano, and John T Lis. *Mol. Ther. Nucleic Acids*, 3(August):e183, 2014. doi: 10.1038/mtna.2014.34.

[65] David R. Latulippe, Kylan Szeto, Abdullah Ozer, Fabiana M. Duarte, Christopher V. Kelly, John M. Pagano, Brian S. White, David Shalloway, John T. Lis, and Harold G. Craighead. *Anal. Chem.*, 85(6):3417–3424, 2013. doi: 10.1021/ac400105e.

[66] H Shi, B E Hoffman, and J T Lis. *Mol. Cell. Biol.*, 17(5):2649–2657, 1997. doi: 10.1128/MCB.17.5.2649.

[67] D. Shangguan, Y. Li, Z. Tang, Z. C. Cao, H. W. Chen, P. Mallikaratchy, K. Sefah, C. J. Yang, and W. Tan. *Proc. Natl. Acad. Sci.*, 103(32):11838–11843, 2006. doi: 10.1073/pnas.0602615103.

[68] Bo Shui, Abdullah Ozer, Warren Zipfel, Nevedita Sahu, Avtar Singh, John T. Lis, Hua Shi, and Michael I. Kotlikoff. *Nucleic Acids Res.*, 40(5), 2012. doi: 10.1093/nar/gkr1264.

[69] Diana Yunusov, Mandy So, Solmaz Shayan, Victor Okhonin, Michael U. Musheev, Maxim V. Berezovski, and Sergey N. Krylov. *Anal. Chim. Acta*, 631(1):102–107, 2009. doi: 10.1016/j.aca.2008.10.027.

[70] Encode Consortion. *Nature*, 489(7414):57–74, 2012. doi: 10.1038/nature11247.

[71] M Kellis, B Wold, M P Snyder, B E Bernstein, A Kundaje, G K Marinov, L D Ward, E Birney, G E Crawford, J Dekker, I Dunham, L L Elnitski, P J Farnham, E A Feingold, M Gerstein, M C Giddings, D M Gilbert, T R Gingeras, E D Green, R Guigo, T Hubbard, J Kent, J D Lieb, R M Myers, M J Pazin, B Ren, J A Stamatoyannopoulos, Z Weng, K P White, and R C Hardison. *Proc Natl Acad Sci U S A*, 111(17):6131–6138, 2014. doi: 10.1073/pnas.1318948111.

[72] P. Nissen. *Science*, 289(5481):920–930, 2000. doi: 10.1126/science.289.5481.920.

[73] Albert E. Dahlberg, R. Brimacombe, J. Atmadja, W. Stiege, D. Schuler, F.H.C. Crick, E. Cundliffe, R. Denman, D. Negre, P.R. Cunningham, K. Nurse, J. Colgan, C. Weitzmann, J. Ofengand, C.C. Hall, D. Johnson, B.S. Cooperman, T.-P. Hausner, J. Atmadja, K.H. Nierhaus, W. Herr, N.M. Chapman, H.F. Noller, A.S. Hui, H.A. DeBoer, A.S. Hui, D.H. Eaton, H.A. de Boer, W.F. Jacob, M. Santer, A.E. Dahlberg, N. Meier, H.U. Goringer, B. Kleuvers, U. Scheibe, J. Eberle, C. Szymkowiak, M. Zacharias, R. Wagner, P. Melancon, C. Lemieux, L. Brakier-Gingras, D. Moazed, H.F. Noller, D. Moazed, H.F. Noller, D. Moazed, H.F. Noller, D. Moazed, B.J. Van Stolk, S. Douthwaite, H.F. Noller, D. Moazed, J.M. Robertson, H.F. Noller, E.J. Murgola, K.A. Hijazi, H.U. Goringer, A.E. Dahlberg, H.F. Noller, M. Asire, A. Barta, S. Douthwaite, T. Goldstein, R. Gutell, D. Moazed, J. Normanly, J.B. Prince, S. Stern, K. Triman, S. Turner, B. Van Stolk, V. Wheaton, B. Weiser, C.R. Woese, H.F. Noller, S. Stern, D. Moazed, T. Powers, P. Svensson, L.-M. Changchien, M. Nomura, W.A. Held, B. Poldermans, H. Bakker, P.H. Van Knippenberg, J.B. Prince, B.H. Taylor, D.L. Thurlow, J. Ofengand, R.A. Zimmermann, M. Santer, C. Sigmund, E. Ettayebi, E.A. Morgan, A.S. Spirin, G. Steiner, E. Kuechler, A. Barta, W.E. Tapprich, W.E. Hill, W.E. Tapprich, D.J. Goss, A.E. Dahlberg, C.L. Thomas, R.J. Gregory, G. Winslow, A. Muto, R.A. Zimmermann, J.F. Thompson, J.E. Hearst, J. Thompson, E. Cundliffe, A.E. Dahlberg, E.N. Trifonov, B. Vester, R.A. Garrett, R.B. Weiss, D.M. Dunn, J.F. Atkins, R.F. Gesteland, R.B. Weiss, D.M. Dunn, A.E. Dahlberg, J.F. Atkins, R.F. Gesteland, C.R. Woese, I Wool, C. Zwieb, D.K. Jemiolo, W.F. Jacob, R. Wagner, and A.E. Dahlberg. *Cell*, 57(4): 525–529, 1989. doi: 10.1016/0092-8674(89)90122-0.

[74] Scott K Silverman. *Org. Biomol. Chem.*, 2:2701–2706, 2004. doi: 10.1039/b411910j.

[75] Scott K. Silverman. In Li Yingfu and Lu Yi, editors, *Funct. Nucleic Acids Anal. Appl.*, number December 2008, pages 47–108. Springer New York, New York, NY, 2009. doi: 10.1007/978-0-387-73711-9_3.

[76] Scott K. Silverman. *Trends Biochem. Sci.*, 41(7):595–609, 2016. doi: 10.1016/j.tibs.2016.04.010.

[77] Miao-chih Tsai, Ohad Manor, Yue Wan, Nima Mosammaparast, Jordon K Wang, Fei Lan, Yang Shi, Eran Segal, and Howard Y Chang. *Science*, 329 (August):689–693, 2010. doi: 10.1126/science.1192002.

[78] Paulo P. Amaral and John S. Mattick. *Mamm. Genome*, 19(7-8):454–492, 2008. doi: 10.1007/s00335-008-9136-7.

[79] Takashi Nagano, Jennifer A Mitchell, Lionel A Sanz, Florian M Pauler, Anne C Ferguson-Smith, Robert Feil, and Peter Fraser. *Science*, 322(5908): 1717–20, 2008. doi: 10.1126/science.1163802.

[80] Radha Raman Pandey, Tanmoy Mondal, Faizaan Mohammad, Stefan Enroth, Lisa Redrup, Jan Komorowski, Takashi Nagano, Debora Mancini-DiNardo, and Chandrasekhar Kanduri. *Mol. Cell*, 32(2):232–246, 2008. doi: 10.1016/j.molcel.2008.08.022.

[81] J. S. Mattick. *EMBO Rep.*, 2(11):986–991, 2001. doi: 10.1093/embo-reports/kve230.

[82] Ra Shivdasani. *Blood*, 108(12):3646–3653, 2006. doi: 10.1182/blood-2006-01-030015.

[83] Me Dinger, Pp Amaral, and Tr Mercer. *Genome Res.*, pages 1433–1445, 2008. doi: 10.1101/gr.078378.108.7.

[84] Pedro J. Batista and Howard Y. Chang. *Cell*, 152(6):1298–1307, 2013. doi: 10.1016/j.cell.2013.02.012.

[85] T R Mercer, M E Dinger, and J S Mattick. *Nat. Rev. Genet.*, 10:155–159, 2009. doi: 10.1038/nrg2521.

[86] Sarah Geisler and Jeff Coller. *Nat. Rev. Mol. Cell Biol.*, 14(11):699–712, 2013. doi: 10.1038/nrm3679.

[87] Naveen K. Navani and Yingfu Li. *Curr. Opin. Chem. Biol.*, 10(3):272–281, 2006. doi: 10.1016/j.cbpa.2006.04.003.

[88] James S Swensen, Yi Xiao, Brian S Ferguson, Arica A Lubin, Rebecca Y Lai, Alan J Heeger, Kevin W Plaxco, and H Tom Soh. *J. Am. Chem. Soc.*, 131(12):4262–4266, apr 2009. doi: 10.1021/ja806531z.

[89] Xiaolei Zuo, Shiping Song, Jiong Zhang, Dun Pan, Lihua Wang, and Chunhai Fan. *J. Am. Chem. Soc.*, 129(5):1042–1043, 2007. doi: 10.1021/ja067024b.

[90] Xiao-Bing Zhang, Rong-Mei Kong, and Yi Lu. *Annu. Rev. Anal. Chem. (Palo Alto. Calif).*, 4:105–28, 2011. doi: 10.1146/annurev.anchem.111808.073617.

[91] Nobuko Hamaguchi, Andrew Ellington, and Martin Stanton. *Anal. Biochem.*, 294(2):126–131, 2001. doi: 10.1006/abio.2001.5169.

[92] Jianwei J Li, Xiaohong Fang, and Weihong Tan. *Biochem. Biophys. Res. Commun.*, 292(1):31–40, 2002. doi: 10.1006/bbrc.2002.6581.

[93] B Scott Ferguson, David A Hoggarth, Dan Maliniak, Kyle Ploense, Ryan J White, Nick Woodward, Kuangwen Hsieh, Andrew J Bonham, Michael Eisenstein, T. E. Kippin, Kevin W Plaxco, and H Tom Soh. *Sci. Transl. Med.*, 5(213):213ra165–213ra165, nov 2013. doi: 10.1126/scitranslmed.3007095.

[94] P. L. Mage, B. S. Ferguson, D. Maliniak, K. L Ploense, T. E. Kippin, and H. T. Soh. *Nat. Biomed. Eng.*, 1(5):0070, may 2017. doi: 10.1038/s41551-017-0070.

[95] Zachary F. Hallberg, Yichi Su, Rebekah Z. Kitto, and Ming C. Hammond. *Annu. Rev. Biochem.*, 86(1):annurev–biochem–060815–014628, jul 2017. doi: 10.1146/annurev-biochem-060815-014628.

[96] Thomas E. Edwards, Daniel J. Klein, and Adrian R. Ferré-D'Amaré. *Curr. Opin. Struct. Biol.*, 17(3):273–279, 2007. doi: 10.1016/j.sbi.2007.05.004.

[97] Jinwei Zhang, Matthew W. Lau, and Adrian R. Ferré-D'Amaré. *Biochemistry*, 49(43):9123–9131, 2010. doi: 10.1021/bi1012645.

[98] Peter Y Watson and Martha J Fedor. *Nat. Chem. Biol.*, 8(12):963–5, 2012. doi: 10.1038/nchembio.1095.

[99] James W. Nelson, Narasimhan Sudarsan, Kazuhiro Furukawa, Zasha Weinberg, Joy X. Wang, and Ronald R. Breaker. *Nat. Chem. Biol.*, 9(12):834–839, oct 2013. doi: 10.1038/nchembio.1363.

[100] J. Bonnet, P. Yin, M. E. Ortiz, P. Subsoontorn, and D. Endy. *Science*, 340 (6132):599–603, may 2013. doi: 10.1126/science.1232758.

[101] J. S. Paige, T. Nguyen-Duc, W. Song, and S. R. Jaffrey. *Science*, 335(6073): 1194–1194, 2012. doi: 10.1126/science.1218298.

[102] Colleen A. Kellenberger, Stephen C. Wilson, Jade Sales-Lee, and Ming C. Hammond. *J. Am. Chem. Soc.*, 135(13):4906–4909, 2013. doi: 10.1021/ja311960g.

[103] Colleen A Kellenberger, Chen Chen, Aaron T Whiteley, Daniel A Portnoy, and Ming C Hammond. *J. Am. Chem. Soc.*, 137(20):6432–5, 2015. doi: 10.1021/jacs.5b00275.

[104] Beatrix Suess, Barbara Fink, Christian Berens, Régis Stentz, and Wolfgang Hillen. *Nucleic Acids Res.*, 32(4):1610–1614, 2004. doi: 10.1093/nar/gkh321.

[105] James M Carothers, Jonathan a Goler, Darmawi Juminaga, and Jay D Keasling. *Science*, 334(6063):1716–1719, dec 2011. doi: 10.1126/science.1212209.

[106] Maung Nyan Win and Christina D Smolke. *Proc. Natl. Acad. Sci. U. S. A.*, 104(36):14283–8, sep 2007. doi: 10.1073/pnas.0703961104.

[107] Dilara Grate and Charles Wilson. *Proc. Natl. Acad. Sci.*, 96(11):6131–6136, 1999. doi: 10.1073/pnas.96.11.6131.

[108] Elena V. Dolgosheina, Sunny C Y Jeng, Shanker Shyam S Panchapakesan, Razvan Cojocaru, Patrick S K Chen, Peter D. Wilson, Nancy Hawkins, Paul A. Wiggins, and Peter J. Unrau. *ACS Chem. Biol.*, 9(10):2412–2420, 2014. doi: 10.1021/cb500499x.

[109] Jeremy S. Paige, Karen Y. Wu, and Samie R. Jaffrey. *Science*, 333(6042): 642–6, 2011. doi: 10.1126/science.1207339.

[110] Grigory S. Filonov, Jared D. Moon, Nina Svensen, and Samie R. Jaffrey. *J. Am. Chem. Soc.*, 136(46):16299–16308, 2014. doi: 10.1021/ja508478x.

[111] Wenjiao Song, Rita L. Strack, Nina Svensen, and Samie R. Jaffrey. *J. Am. Chem. Soc.*, 136(4):1198–1201, 2014. doi: 10.1021/ja410819x.

[112] Georgios Pothoulakis, Francesca Ceroni, Benjamin Reeve, and Tom Ellis. *ACS Synth. Biol.*, 3(3):182–187, 2014. doi: 10.1021/sb400089c.

[113] O Shimomura, M Chalfie, and R Tsien. *Back to cited text*, 2008.

[114] Daniel H. Appella. *Curr. Opin. Chem. Biol.*, 13(5-6):687–696, 2009. doi: 10.1016/j.cbpa.2009.09.030.

[115] Lucile Crouzier, Camille Dubois, Stacey L Edwards, Lasse H Lauridsen, Jesper Wengel, and Rakesh N Veedu. *PLoS One*, 7(4):e35990, 2012. doi: 10.1371/journal.pone.0035990.

[116] R. Padilla. *Nucleic Acids Res.*, 30(24):138e–138, 2002. doi: 10.1093/nar/gnf138.

[117] Mark A. Behlke. *Oligonucleotides*, 18(4):305–320, 2008. doi: 10.1089/oli.2008.0164.

[118] Doug Irvine, Craig Tuerk, and Larry Gold. *J. Mol. Biol.*, 222(3):739–761, 1991. doi: 10.1016/0022-2836(91)90509-5.

[119] Jinpeng Wang, Joseph F. Rudzinski, Qiang Gong, H. Tom Soh, and Paul J. Atzberger. *PLoS One*, 7(8):1–8, 2012. doi: 10.1371/journal.pone.0043940.

[120] Barry Vant-Hull, Antonio Payano-Baez, Robert H Davis, and Larry Gold. *J. Mol. Biol.*, 278(3):579–597, 1998. doi: 10.1006/jmbi.1998.1727.

[121] Marko Djordjevic. *Biomol. Eng.*, 24(2):179–189, 2007. doi: 10.1016/j.bioeng.2007.03.001.

[122] Jonghoon Kang, Soog Lee Myung, and David G. Gorenstein. *J. Biochem. Biophys. Methods*, 64:147–151, 2005. doi: 10.1016/j.jbbm.2005.06.003.

[123] M. N. Stojanovic, P. de Prada, and D. W. Landry. *J. Am. Chem. Soc.*, 123 (21):4928–4931, 2001. doi: 10.1021/ja0038171.

[124] Alexis Vallée-Bélisle and Kevin W. Plaxco. *Curr. Opin. Struct. Biol.*, 20(4): 518–526, 2010. doi: 10.1016/j.sbi.2010.05.001.

[125] Eugene W M Ng, David T Shima, Perry Calias, Emmett T Cunningham, David R Guyer, and Anthony P Adamis. *Nat. Rev. Drug Discov.*, 5(2): 123–32, 2006. doi: 10.1038/nrd1955.

[126] Milan N. Stojanovic and Dmitry M. Kolpashchikov. *J. Am. Chem. Soc.*, 126 (30):9266–9270, 2004. doi: 10.1021/ja032013t.

[127] Sanjay Tyagi and F R Kramer. *Nat. Biotechnol.*, 14(3):303–308, 1996. doi: 10.1038/nbt0696-765.

[128] S Jhaveri, M Rajendran, and a D Ellington. *Nat. Biotechnol.*, 18(12): 1293–1297, 2000. doi: 10.1038/82414.

[129] S. S. Oh, K. Plakos, X. Lou, Y. Xiao, and H. T. Soh. *Proc. Natl. Acad. Sci.*, 107(32):14053–14058, 2010. doi: 10.1073/pnas.1009172107.

[130] Kareem M. Ahmad, Yi Xiao, and H. Tom Soh. *Nucleic Acids Res.*, 40(22): 11777–11783, 2012. doi: 10.1093/nar/gks899.

[131] Jan Hoinka, Elena Zotenko, Adam Friedman, Zuben E. Sauna, and Teresa M. Przytycka. *Bioinformatics*, 28(12):215–223, 2012. doi: 10.1093/bioinformatics/bts210.

[132] Jan Hoinka, Alexey Berezhnoy, Zuben E. Sauna, Eli Gilboa, and Teresa M. Przytycka. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 8394 LNBI:115–128, 2014. doi: 10.1007/978-3-319-05269-4_9.

[133] Khalid K Alam, Jonathan L Chang, and Donald H Burke. *Mol. Ther. Acids*, 4(August 2014):e230, 2015. doi: 10.1038/mtna.2015.4.

[134] Minseon Cho, Seung Soo Oh, Jeff Nie, Ron Stewart, Michael Eisenstein, James Chambers, Jamey D Marth, Faye Walker, James a Thomson, and H Tom Soh. *Proc. Natl. Acad. Sci. U. S. A.*, 110(46):18460–5, 2013. doi: 10.1073/pnas.1315866110.

[135] Nicholas O. Fischer, Jeffrey B H Tok, and Theodore M. Tarasow. *PLoS One*, 3(7):1–9, 2008. doi: 10.1371/journal.pone.0002720.

[136] Razvan Nutiu, Robin C Friedman, Shujun Luo, Irina Khrebtukova, David Silva, Robin Li, Lu Zhang, Gary P Schroth, and Christopher B Burge. *Nat. Biotechnol.*, 29(7):659–664, 2011. doi: 10.1038/nbt.1882.

[137] Jason D Buenrostro, Carlos L Araya, Lauren M Chircus, Curtis J Layton, Howard Y Chang, Michael P Snyder, and William J Greenleaf. *Nat. Biotechnol.*, 32(6):562–8, 2014. doi: 10.1038/nbt.2880.

[138] Jacob M Tome, Abdullah Ozer, John M Pagano, Dan Gheba, Gary P Schroth, and John T Lis. *Nat. Methods*, 11(6):683–688, may 2014. doi: 10.1038/nmeth.2970.

[139] Walter Gilbert. *Nature*, 319(6055):618–618, feb 1986. doi: 10.1038/319618a0.

[140] Christopher G. Knight, Mark Platt, William Rowe, David C. Wedge, Farid Khan, Philip J R Day, Andy Mcshea, Joshua Knowles, and Douglas B. Kell. *Nucleic Acids Res.*, 37(1):1–10, 2009. doi: 10.1093/nar/gkn899.

[141] Rita L Strack, Matthew D Disney, and Samie R Jaffrey. *Nat. Methods*, 10 (12):1219–1224, 2013. doi: 10.1038/nmeth.2701.

[142] Maumita Mandal, Benjamin Boese, Jeffrey E. Barrick, Wade C. Winkler, and Ronald R. Breaker. *Cell*, 113(5):577–586, 2003. doi: 10.1016/S0092-8674(03)00391-X.

[143] J. A. Cruz, M.-F. Blanchet, Michal Boniecki, J. M. Bujnicki, S.-J. Chen, S. Cao, R. Das, F. Ding, N. V. Dokholyan, S. C. Flores, L. Huang, C. A.

Lavender, V. Lisi, F. Major, K. Mikolajczak, D. J. Patel, A. Philips, T. Puton, J. Santalucia, F. Sijenyi, T. Hermann, K. Rother, M. Rother, A. Serganov, M. Skorupski, T. Soltysinski, P. Sripakdeevong, I. Tuszynska, K. M. Weeks, C. Waldsich, M. Wildauer, N. B. Leontis, and E. Westhof. *RNA*, 18(4): 610–625, apr 2012. doi: 10.1261/rna.031054.111.

[144] Zhichao Miao, Ryszard W. Adamiak, Marc-Frédérick Blanchet, Michal Boniecki, Janusz M. Bujnicki, Shi-Jie Chen, Clarence Cheng, Grzegorz Chojnowski, Fang-Chieh Chou, Pablo Cordero, José Almeida Cruz, Adrian R. Ferré-D'Amaré, Rhiju Das, Feng Ding, Nikolay V. Dokholyan, Stanislaw Dunin-Horkawicz, Wipapat Kladwang, Andrey Krokhotin, Grzegorz Lach, Marcin Magnus, François Major, Thomas H. Mann, Benoît Masquida, Dorota Matelska, Mélanie Meyer, Alla Peselis, Mariusz Popenda, Katarzyna J. Purzycka, Alexander Serganov, Juliusz Stasiewicz, Marta Szachniuk, Arpit Tandon, Siqi Tian, Jian Wang, Yi Xiao, Xiaojun Xu, Jinwei Zhang, Peinan Zhao, Tomasz Zok, and Eric Westhof. *Rna*, 21(6):1066–1084, 2015. doi: 10.1261/rna.049502.114.

[145] Zhichao Miao, Ryszard W. Adamiak, Maciej Antczak, Robert T. Batey, Alexander J. Becka, Marcin Biesiada, Michał J. Boniecki, Janusz Bujnicki, Shi-Jie Chen, Clarence Yu Cheng, Fang-Chieh Chou, Adrian R. Ferré-D'Amaré, Rhiju Das, Wayne K. Dawson, Ding Feng, Nikolay V. Dokholyan, Stanisław Dunin-Horkawicz, Caleb Geniesse, Kalli Kappel, Wipapat Kladwang, Andrey Krokhotin, Grzegorz E. Łach, François Major, Thomas H. Mann, Marcin Magnus, Katarzyna Pachulska-Wieczorek, Dinshaw J. Patel, Joseph A. Piccirilli, Mariusz Popenda, Katarzyna J. Purzycka, Aiming Ren, Greggory M. Rice, John Santalucia, Joanna Sarzynska, Marta Szachniuk, Arpit Tandon, Jeremiah J. Trausch, Siqi Tian, Jian Wang, Kevin M. Weeks, Benfeard Williams, Yi Xiao, Xiaojun Xu, Dong Zhang, Tomasz Zok, and Eric Westhof. *Rna*, page rna.060368.116, 2017. doi: 10.1261/rna.060368.116.

[146] Daniel M. Held, S. Travis Greathouse, Amit Agrawal, and Donald H. Burke. *J. Mol. Evol.*, 57(3):299–308, 2003. doi: 10.1007/s00239-003-2481-y.

[147] Zhen Huang and Jack W Szostak. *RNA*, 9:1456–1463, 2003. doi: 10.1261/rna.5990203.

[148] Patrick C. Cirino, Kimberly M. Mayer, and Daisuke Umeno. In *Dir. Evol. Libr. Creat.*, volume 231, pages 3–10. Humana Press, New Jersey, 2003. doi: 10.1385/1-59259-395-X:3.

[149] Byounghoon Hwang and Seong-Wook Lee. *Biochem. Biophys. Res. Commun.*, 290(2):656–662, 2002. doi: 10.1006/bbrc.2001.6252.

[150] Joshua A. Bittker, Brian V. Le, and David R. Liu. *Nat. Biotechnol.*, 20(10): 1024–1029, 2002. doi: 10.1038/nbt736.

[151] Bradley Hall, John M. Micheletti, Pooja Satya, Krystal Ogle, Jack Pollard, and Andrew D. Ellington. In *Curr. Protoc. Mol. Biol.*, number SUPPL. 88, pages 1–27. John Wiley & Sons, Inc., Hoboken, NJ, USA, oct 2009. doi: 10.1002/0471142727.mb2402s88.

[152] Jonathan H Davis and Jack W Szostak. *Proc. Natl. Acad. Sci. U. S. A.*, 99 (18):11616–11621, 2002. doi: 10.1073/pnas.182095699.

[153] Jinpeng Wang, Qiang Gong, Nupur Maheshwari, Michael Eisenstein, Mary Luz Arcila, Kenneth S. Kosik, and H. Tom Soh. *Angew. Chemie - Int. Ed.*, 53:4796–4801, 2014. doi: 10.1002/anie.201309334.

[154] Jinpeng Wang, Jingwen Yu, Qin Yang, John McDermott, Alexander Scott, Matthew Vukovich, Remy Lagrois, Qiang Gong, William Greenleaf, Michael Eisenstein, B. Scott Ferguson, and H. Tom Soh. *Angew. Chemie - Int. Ed.*, 56(3):744–747, 2017. doi: 10.1002/anie.201608880.

[155] Ignacio Tinoco and Carlos Bustamante. *J. Mol. Biol.*, 293(2):271–281, 1999. doi: 10.1006/jmbi.1999.3001.

[156] Martin Chalfie and Biological Sciences. *Photochem. Photobiol.*, 62(4):651–656, 1995. doi: 10.1111/j.1751-1097.1995.tb08712.x.

[157] Karin Nienhaus and G. Ulrich Nienhaus. *Chem. Soc. Rev.*, 43(4):1088–1106, 2014. doi: 10.1039/C3CS60171D.

[158] Tsai-Wen Chen, Trevor J. Wardill, Yi Sun, Stefan R. Pulver, Sabine L. Renninger, Amy Baohan, Eric R. Schreiter, Rex A. Kerr, Michael B. Orger, Vivek Jayaraman, Loren L. Looger, Karel Svoboda, and Douglas S. Kim. *Nature*, 499(7458):295–300, 2013. doi: 10.1038/nature12354.

[159] G N Phillips. *Curr. Opin. Struct. Biol.*, 7(6):821–827, 1997. doi: Doi10.1016/S0959-440x(97)80153-4.

[160] Jae Hyun Bae, Marina Rubini, Gregor Jung, Georg Wiegand, Markus H J Seifert, M. Kamran Azim, Jeong Sun Kim, Andreas Zumbusch, Tad A. Holak, Luis Moroder, Robert Huber, and Nediljko Budisa. *J. Mol. Biol.*, 328(5): 1071–1081, 2003. doi: 10.1016/S0022-2836(03)00364-4.

[161] Mingxu You, Jacob L. Litke, and Samie R. Jaffrey. *Proc. Natl. Acad. Sci.*, 112(21):E2756–E2765, 2015. doi: 10.1073/pnas.1504354112.

[162] Floyd J. Green. *The Sigma-Aldrich Handbook of Stains, Dyes and Indicators*. Aldrich Chemical Company, Inc, Milwaukee, Wisconsin, 1990.

[163] Fabio M. Spiga, Paolo Maietta, and Carlotta Guiducci. *ACS Comb. Sci.*, 17 (5):326–333, 2015. doi: 10.1021/acscombsci.5b00023.

[164] Tatjana Schütze, Barbara Wilhelm, Nicole Greiner, Hannsjörg Braun, Franziska Peter, Mario Mörl, Volker a. Erdmann, Hans Lehrach, Zoltán Konthur, Marcus Menger, Peter F. Arndt, and Jörn Glökler. *PLoS One*, 6 (12):1–10, 2011. doi: 10.1371/journal.pone.0029604.

[165] D Herschlag. *J. Biol. Chem.*, 270:20871–20874, 1995. doi: 10.1074/jbc.270.36.20871.

[166] Andreas R. Gruber, Ronny Lorenz, Stephan H. Bernhart, Richard Neuböck, and Ivo L. Hofacker. *Nucleic Acids Res.*, 36(Web Server issue):70–74, 2008. doi: 10.1093/nar/gkn188.

[167] Ronny Lorenz, Stephan H Bernhart, Christian Höner zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. *Algorithms Mol. Biol.*, 6(1):26, 2011. doi: 10.1186/1748-7188-6-26.

[168] J. Kanhere, R. Gopinathan, and J. Banerjee. *Water. Air. Soil Pollut.*, 225 (9), 2014. doi: 10.1007/s11270-014-2134-3.

[169] Annalaura Stammati, Carlo Nebbia, Isabella De Angelis, Alessandra Giuliano Albo, Monica Carletti, Claudia Rebecchi, Franco Zampaglioni, and Mauro Dacasto. *Toxicol. Vitr.*, 19(7):853–858, 2005. doi: 10.1016/j.tiv.2005.06.021.

[170] Sandra Culp and Frederick Beland. *Int. J. Toxicol.*, 15(3):219–238, 1996. doi: 10.3109/10915819609008715.

[171] Saumya Saurabh, Adam M. Perez, Colin J. Comerci, Lucy Shapiro, and W. E. Moerner. *J. Am. Chem. Soc.*, 138(33):10398–10401, 2016. doi: 10.1021/jacs.6b05943.

[172] Luc Ponchon and Frédéric Dardel. *Nat. Methods*, 4(7):571–576, 2007. doi: 10.1038/nmeth1058.

[173] GrigoryS. Filonov, ChristinaW. Kam, Wenjiao Song, and SamieR. Jaffrey. *Chem. Biol.*, 22(5):649–660, 2015. doi: 10.1016/j.chembiol.2015.04.018.

[174] P D Good, A J Krikos, S X L Li, E Bertrand, N S Lee, L Giver, A Ellington, J A Zaia, J J Rossi, and D R Engelke. *Gene Ther.*, 4(1):45–54, 1997. doi: 10.1038/sj.gt.3300354.

[175] Christopher Baugh, Dilârâ Grate, and Charles Wilson. *J. Mol. Biol.*, 301 (1):117–128, 2000. doi: 10.1006/jmbi.2000.3951.

[176] Didier Piau. *J. Comput. Biol.*, 9(6):831–47, 2002. doi: 10.1089/10665270260518308.

[177] Georgi K. Marinov, Brian A. Williams, Ken McCue, Gary P. Schroth, Jason Gertz, Richard M. Myers, and Barbara J. Wold. *Genome Res.*, 24(3):496–510, 2014. doi: 10.1101/gr.161034.113.

[178] Saiful Islam, Una Kjällquist, Annalena Moliner, Pawel Zajac, Jian Bing Fan, Peter Lönnerberg, and Sten Linnarsson. *Genome Res.*, 21(7):1160–1167, 2011. doi: 10.1101/gr.110882.110.

[179] John S. Mattick and Igor V. Makunin. *Hum. Mol. Genet.*, 15 Spec No(1): 17–29, 2006. doi: 10.1093/hmg/ddl046.

[180] Pascale Romby and Mathias Springer. *Trends Genet.*, 19(3):155–161, 2003. doi: 10.1016/S0168-9525(03)00020-9.

[181] Jörg Stülke. *Arch. Microbiol.*, 177(6):433–440, 2002. doi: 10.1007/s00203-002-0407-5.

[182] A. Wachter, M. Tunc-Ozdemir, B. C. Grove, P. J. Green, D. K. Shintani, and R. R. Breaker. *Plant Cell Online*, 19(11):3437–3450, 2007. doi: 10.1105/tpc.107.053645.

[183] Heinz W. Pley, Kevin M. Flaherty, and David B. McKay. *Nature*, 372(6501): 68–74, 1994. doi: 10.1038/372068a0.

[184] N. Sudarsan, M. C. Hammond, K. F. Block, R. Welz, J. E. Barrick, A. Roth, and R. R. Breaker. *Science*, 314(5797):300–304, oct 2006. doi: 10.1126/science.1130716.

[185] Hao Qu, Andrew T. Csordas, Jinpeng Wang, Seung Soo Oh, Michael S. Eisenstein, and Hyongsok Tom Soh. *ACS Nano*, 10(8):7558–7565, 2016. doi: 10.1021/acsnano.6b02558.

[186] Marat D Kazanov, Alexey G Vitreschak, and Mikhail S Gelfand. *BMC Genomics*, 8(1):347, 2007. doi: 10.1186/1471-2164-8-347.

[187] Pablo Ceres, Andrew D. Garst, Joan G. Marcano-Velázquez, and Robert T. Batey. *ACS Synth. Biol.*, 2(8):463–472, 2013. doi: 10.1021/sb4000096.

[188] Brian J. Tucker and Ronald R. Breaker. *Curr. Opin. Struct. Biol.*, 15(3 SPEC. ISS.):342–348, 2005. doi: 10.1016/j.sbi.2005.05.003.

[189] Yoko Nomura and Yohei Yokobayashi. *J. Am. Chem. Soc.*, 129(45):13814–13815, 2007. doi: 10.1021/ja076298b.

[190] Shawn K. Desai and Justin P. Gallivan. *J. Am. Chem. Soc.*, 126(41): 13247–13254, 2004. doi: 10.1021/ja048634j.

[191] Shana Topp and Justin P. Gallivan. *ChemBioChem*, 9(2):210–213, 2008. doi: 10.1002/cbic.200700546.

[192] Sean A. Lynch and Justin P. Gallivan. *Nucleic Acids Res.*, 37(1):184–192, 2009. doi: 10.1093/nar/gkn924.

[193] Joe C. Liang, Andrew L. Chang, Andrew B. Kennedy, and Christina D. Smolke. *Nucleic Acids Res.*, 40(20):1–14, 2012. doi: 10.1093/nar/gks636.

[194] Alexander Wittmann and Beatrix Suess. *Mol. Biosyst.*, 7(8):2419, 2011. doi: 10.1039/c1mb05070b.

[195] Kathy Y. Wei, Yvonne Y. Chen, and Christina D. Smolke. *Biotechnol. Bioeng.*, 110(4):1201–1210, 2013. doi: 10.1002/bit.24792.

[196] Xin C. Wang, Stephen C. Wilson, and Ming C. Hammond. *Nucleic Acids Res.*, 44(17):1–10, 2016. doi: 10.1093/nar/gkw580.

[197] Yichi Su, Scott F. Hickey, Samantha G L Keyser, and Ming C. Hammond. *J. Am. Chem. Soc.*, 138(22):7040–7047, 2016. doi: 10.1021/jacs.6b01621.